STRUCTURE DETERMINATION OF PROTEINS AND PROTEIN AGGREGATES BY MAGIC-ANGLE SPINNING SOLID-STATE NMR

BY

ANDREW JON NIEUWKOOP

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Associate Professor Chad M. Rienstra, Chair
Professor Eric Oldfield
Professor Klaus Schulten
Assistant Professor Julia M. George

ABSTRACT

Solid state NMR (SSNMR) is a structure determination technique uniquely suited to study protein aggregates and fibrils. Unlike solution NMR or X-ray crystallography, SSNMR can obtain atomic resolution structural information on samples of protein fibrils which are insoluble and do not produce X-ray diffracting crystals. As SSNMR begins to realize this potential, new structure determination techniques will be important in enable SSNMR to investigate ever larger and more complicated systems.

Alpha-synuclein (AS) is the primary protein component of Lewy bodies, the pathological hallmark of Parkinson's disease. The structure of AS in its fibril form is unknown, as is the mechanism by which it contributes to neurodegeneration. At 140 residues, AS is much larger than other fibril systems that have been studied by SSNMR. To solve the structure of AS fibrils will require combining new pulse sequences with advanced isotopic labeling schemes, and novel structure calculation methods. The techniques developed in this study will be useful in the study of other protein aggregates, as well as membrane proteins and complexes, for which SSNMR is the structure determination method of choice.

For my Mom and Dad

Acknowledgements

I'd like to thank my advisor Professor Chad Rienstra for his guidance and for allowing me the opportunity to work with some of the best solid-state NMR spectrometers in the world. His personal instruction at those spectrometers, working with the data and throughout the analysis, has been invaluable to me. I'd like to thank many former members of the Rienstra group: Dr. Ben Wylie for his tutelage at the spectrometer, for teaching me the inner workings of MINUIT and for showing me how to use the python extensions for XPLOR; Dr. Trent Franks for teaching me to pulse code and how to run XPLOR; Dr. Donghua Zhou for teaching me all about his proton detected experiments, for answering all my questions about them after he'd left, and for every beautiful awk script he left the group; and Dr. Kathy Kloepper for her leadership of the AS project and for developing the infrastructure for producing the many AS samples I've used over the years. Dr. Lindsay Sperling was a stalwart lab mate from that first July day when we got to UIUC and promptly moved office furniture. From doing homework, practicing Lit Seminars, Prelims and Post Doc talks, to games of poker and pool, I couldn't have done this without her.

I'd like to thank the entire Rienstra group for their support over the years: Dr. Deb Berthold and Luisel Lemkau for the innumerable samples they made for me over the years, and for putting up with me when I flailed about the wet lab; Every undergrad past and present who helped make samples, Shin Lee Wook, Lars Rikardsen, Elliot Brea, Gautam Shah, Reika Ebisu, Daniel Ladror and Kevin Hartman.

I'd like to thank everyone at the NMR Facility at the University for their assistance at the spectrometer and for all the things they do behind the scenes to make what we do

possible: Dr. Vera Mainz, Dr. Dean Olson, Dr. Jen Rapp, Dr. Feng Li, Dr. Trent Franks, Tracie Hubert and Ben Fisher.

I'd like to thank Mike Hallock for his patience in teaching me everything I know about Linux and system administration, and for dealing with all the problems I couldn't figure out and/or caused with our computer systems over the years.

I'd like to thank all my professors at Carleton College, for giving me the basis from which to tackle the world: specifically, Dr. Joe Chihade for teaching me what real scientific research is, and how to do it, Dr. Dani Kohen for showing me what computers can do with chemistry and for convincing me to apply to Illinois, and Dr. Dave Alberg, for introducing me to NMR and the best one legged demonstration of multiple processing frequencies ever performed.

I'd like to thank Gus Lukow at West Ottawa High School for the best A.P. Chem class that has ever been taught. You ignited my love of chemistry that year, and it has shaped my life ever since.

I'd like to thank my friends and family for their love and support throughout this experience called grad school. Without your wisdom, advice and perspective, I would certainly have been overwhelmed. Most especially my parents, who have given me every opportunity a child could ask for, whose example I will always try to follow, and whose continuing love and friendship I could not do without.

Table of Contents

List of Figures

List of Tables

Chapter 1

Methods for Protein Structure Determination by Magic-Angle Spinning Solid State NMR

1.1 Introduction

Atomic resolution structures are the cornerstone of the modern study of protein function, providing a detailed view of how a protein performs the chemistry it evolved to complete. For years the gold standard for protein structure determination has been X-ray crystallography, accounting for almost 90% of the ~70,000 protein structures in the protein data bank (www.pdb.org). In the 1990's, solution NMR was developed as an additional tool for protein structure determination. Solid-state NMR (SSNMR), in particular magic angle spinning (MAS) SSNMR, offers the potential to apply many of the techniques developed for solution NMR to systems not amenable to that technique. Examples of systems for which SSNMR is uniquely suited—due to lack of solubility or a tendency not to form diffracting crystals—include insoluble fibrils, protein aggregates and membrane proteins. Now with improvements in instrumentation, such as higher field magnets, more stable high power amplifiers and fast magic angle spinning probes, the potential of SSNMR is being realized and several research groups have now solved protein structures to atomic resolution.[1-11]

Aligned sample static SSNMR techniques complement MAS methods. For example, Cross and Opella have spearheaded such efforts and determined the membrane orientation of the transmembrane helices of several membrane peptides. The techniques and hardware used in these studies are somewhat different from those

utilized in MAS SSNMR so they will not be covered in this introduction. Rather, we refer the reader to reviews from Opella that summarize the capabilities of this technique.[12, 13]

1.2 Nanocrystalline Proteins

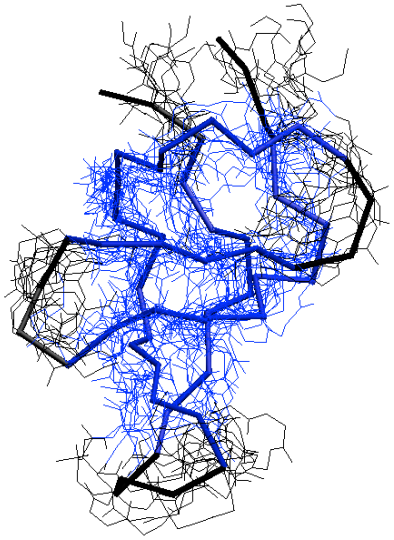

Figure 1.1: The structure of nanocrystalline SH3 solved by Oschkinat and co-workers.

Twelve of the 15 lowest energy structures (PDB ID: 1M8M) out of 200 calculated. Blue segments indicate areas of beta-sheet, black loops and turns. One structure is displayed with thicker lines to guide the eye.

Similar to the progression of solution NMR structures,[14, 15] the first structures solved by SSNMR followed a pattern of qualitative heteronuclear distance and semi-empirical backbone dihedral angle restraints. These early structures primarily relied on relatively simple ^{13}C-detected experiments, such as 2D ^{13}C-^{13}C correlation utilizing either DARR[16] or PDSD[17] mixing. These through space correlations served as distance constraints in simulated annealing protein structure determination calculations using programs such as XPLOR[18] and CNS.[19] Dihedral angels were determined using the empirical TALOS

method.[20] In this way, the structures of several small proteins were solved to moderate resolution,[1-3] providing experimental benchmarks for extensively isotopically labeled (^{13}C, ^{15}N) samples. Figure 1.1 shows the structure of SH3 solved by Oschkinat and co-workers (PDB ID: 1M8M).[1] To improve sensitivity and resolution, in order to resolve long-range distances, this study utilized fractional 1,3-^{13}C- and 2-^{13}C-glycerol labeling, setting an important precedent for subsequent structural studies. This method developed by LeMaster,[21] and demonstrated in solids by Hong,[22] produces a checker board pattern of ^{13}C enrichment. This greatly attenuates ^{13}C-^{13}C scalar couplings, and alleviates spectral degeneracy by decreasing the number of labeled sites. As such, it has been a key development to enable much of the ^{13}C protein work discussed here.

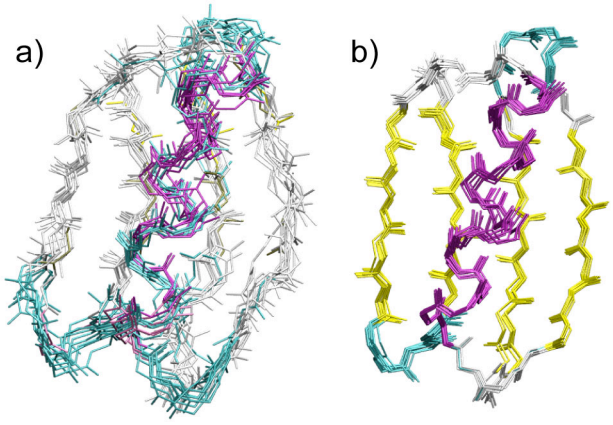

Figure 1.2: SSNMR structure of GB1 solved by Franks et al.

Structure calculated (a) with distances only and (b) distances, dihedrals and VEAN restraints. The structure with all the restraints included (b, PDB ID: 2JSV) has a backbone RMSD of 0.3 Å.

Once it had been shown that protein structures could be solved, the next major achievement was to solve these structures to atomic resolution. Several strategies have been employed to increase the precision of NMR structures, by increasing the quantities of restraints and/or improving the precision of angular and/or distance determinations.

Franks et al. demonstrated in 2008 that by using a substantially larger number, thousands as opposed to hundreds, of qualitative homonuclear correlations from ^{13}C-^{13}C glycerol labeled spectra, it was possible to solve the structure of GB1 to ~1 Å bbRMSD.[7] However, much greater improvement was realized by the use of more precise VEctor ANgle (VEAN) restraints,[23] enabling a structure of GB1 with a backbone RMSD of 0.3 Å. Figure 1.2 shows the progression from (a) distances only to (b) distances with TALOS and VEANs. Nieuwkoop et al. demonstrated that highly precise distances from 3D heteronuclear TEDOR experiments (see Chapter 2)[24] and backbone conformations from chemical shift tensors[25, 26] can be used to solve similarly high resolution protein structures. The remaining challenge is to apply these highly precise techniques to larger protein systems and/or novel topologies that are observed in membrane proteins and/or oligomeric assemblies.

1.3 Protein Aggregates and Fibrils

MAS-SSNMR has made a number of unique contributions to the structural study of protein aggregates and fibrils.[27-32] These aggregates are far too large (>MDa) for solution NMR, essentially rendering their correlation times equivalent to insoluble particles, even in cases where the fibrils can be solubilized. Fibrils do not contain the long-range 3D order needed to diffract x-rays to high resolution. Initial SSNMR efforts focused on small, fibril forming peptides, which could be produced via solid phase peptide synthesis. This strategy enabled site-specific incorporation of labeled amino acids, simplifying the assignment procedure. Much of the work on Aβ, spearheaded by Tycko and coworkers, utilized samples labeled at a few sites at a time in order to determine assignments and detect long-range correlations. These were used to build a

model of the fold of Aβ40.[33] Illustrating the challenges posed by such systems, this model was expanded upon several times,[30, 34, 35] and structures of different states determined to have subtle characteristic features, e.g., a change from two-fold to three-fold symmetry. Jaroniec et al. solved the structure of the monomer unit of fibrils of the 12 residue peptide transthyretin in 2004,[36] work which has just been placed into larger context recently with the determination of the quaternary structure of the fibirls.[37]

A major step forward for the field was made by the Meier group in 2008 when they solved the atomic resolution structure of HET-s fibrils.[9] The fold of these fibrils was unknown before their study, making this the first protein structure solved by MAS-SSNMR. Two years later, they released a refined version of the structure, along with a detailed discussion of the experiments and samples needed to unambiguously determine such a structure.[38] The principal consideration is the need for discrimination between inter- and intra-molecular contacts. This is achieved through the use of diluted samples, where ^{13}C labeled protein is diluted in ^{13}C-depleted protein. In such a sample any ^{13}C-^{13}C contacts observed must be intramolecular, while peaks that do not appear in the diluted sample but do in a uniformly labeled sample are assumed to be intermolecular. In addition to HET-s, this strategy was used to remove the intermolecular peaks in many monomer structure calculations, while site-specific carbonyl labels were utilized in the transthyretin study. A complementary strategy involves placing ^{13}C labels on one molecule and ^{15}N labels on another, then detecting intermolecular contacts through the use of a heteronuclear correlation sequence such as PAIN-CP,[39] NHHC[40] or TEDOR.[41, 42] This strategy has been used to observe the

interfaces of the homodimer Crh,[8] reassembled thioredoxin,[43] the multimer αB crystalline[11] and nanocrystals of GB1.[44]

1.4 Membrane Proteins

In addition to work on protein fibrils, SSNMR has also made great strides in solving membrane protein structures in the native bilayer environment. Membrane proteins, particularly when associated with lipids, are notoriously difficult to crystallize. Recent technologies such as lipidic cubic phase crystallogenesis[45] have increased the likelihood of success, resulting in landmark achievements.[46, 47] Nevertheless, there are relatively few X-ray structures of membrane proteins, and the structures that are available have required Herculean efforts involving hundreds or thousands of potential targets and decades of work.[48] In solution NMR, smaller membrane proteins in detergent micelles remain within the size limit for effective structural studies. SSNMR on the other hand has the advantage of no theoretical inherent upper size limit, which opens the potential for more physiological membrane environments and larger membrane systems. Recent studies of proteorhodopsin,[49] bacteriorhodopsin[50, 51] and DsbB[52, 53] have demonstrated that SSNMR assignment strategies–3D and 4D pulse sequences, with uniform or glycerol ^{13}C labeling–are effective on membrane proteins as well. Hong and co-workers have shown that the structures of antimicrobial peptides[54, 55] and small proteins[10] can be solved through standard structure determination methods, coupled with ^{2}H orientation data specifically helpful in membrane contexts. The van Rossum group has solved the structure of the trimeric membrane protein YadA (Experimental NMR Conference, Asilomar, CA, April 2011), using the now-standard SSNMR methods of partial and glycerol labeling. This study illustrates that such

methods are effective for membrane protein structural studies and represent yet another step forward in the field SSNMR structure determination. SSNMR restraints in combination with X-ray data can produce structures better than either set of data alone. For example, SSNMR data were used to refine the low resolution X-ray structure of the membrane protein DsbB,[56] and SSNMR data combined with small angle X-ray scattering data were used to solve the structure of αB crystallin.[11] In both cases yielding additional information compared to what the X-ray data alone can show.

For higher molecular weight membrane protein complexes, SSNMR offers the ability to provide chemical details at atomic resolution. In the example of the E-coli cytochrome bo_3 oxidase, at 144 kDa membrane protein complex, the crystal structure lacks prosthetic groups and therefore the chemical detail needed to evaluate specific mechanistic hypotheses. However, SSNMR has shown that even in such large systems individual sites can be resolved and that specific labeling schemes will enable structural information to be determined in the active site of the large protein.[57]

1.5 Proton Detection

As solid-state NMR pushes towards larger and more complicated structures, the sensitivity and resolution increasingly become a limiting factor. However, with advances in the MAS technology, MAS rates of 40 kHz and higher are now obtainable, sufficient to attenuate ^{1}H-^{1}H dipolar couplings and provide ~0.1 to 0.2 ppm linewidths, enabling the benefits of ^{1}H-detected SSNMR to be realized. Due to the high gyromagnetic ratio and natural abundance of ^{1}H, sensitivity can be improved by a factor of ~8 to 30 relative to ^{13}C or ^{15}N detection,[58] and have been shown to be applicable to solid systems by the Reif and Zilm groups.[59-62] The theoretical sensitivity enhancement over ^{15}N and ^{13}C

detection, considering just the gyromagnetic ratio, would be ~30 and 8 respectively, the ratio of gamma to the 3/2 power.[63]

In addition to the fast spinning modules, suppression of solvent ^{1}H signals is also vital to realize the full dynamic range of ^{1}H signals. The MISSISSIPPI scheme has proven robust at suppression in the range of 100 to 1,000-fold suppression, and also effectively suppresses signals at multiple frequencies, thus eliminating the signals from non-aqueous solvents or precipitation reagents, such as methylpentanediol and 2-propanol used for GB1 samples. The MISSISSIPPI scheme is even effective for background signals with short ^{1}H longitudinal relaxation times.[64]

The dilution of protons through deuteration of protein samples facilitates further improvements in resolution and sensitivity.[65] In order to detect ^{1}H spins, back-exchange of labile sites of the protein after expression can be used. The level of ^{1}H back-exchange which leads to the best overall signal intensity has been investigated by the Oschkinat group, and was determined to be ~35% for nanocrystals of SH3.[66] This result is likely to be general for microcrystalline proteins, but requires further investigation in the context of membrane proteins and fibrils.

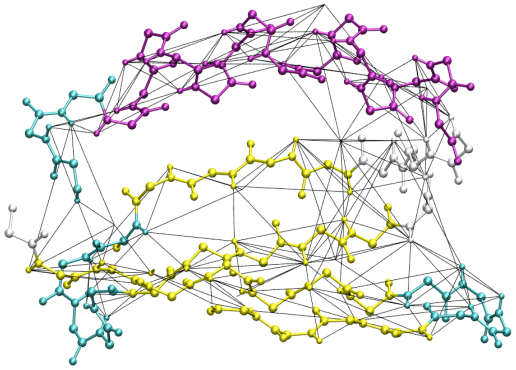

Figure 1.3: Structure of GB1 solved using proton detected ^{1}H-^{1}H correlations.

Structure of GB1 displayed in CPK mode colored by secondary structure. ^{1}H-^{1}H distances used in the calculation are shown on the lowest energy structure.

With the experimental methods developed to the point 3D ^{1}H detected experiments were possible, the Rienstra group utilized a suite of 3D experiments to assign the HN protons of GB1, obtain long-range ^{1}H-^{1}H distances and solve the structure of GB1 (Figure 1.3, PDB 1D: 2JU6).[67] Capitalizing on the work of Kay and others[68] in the use of α-keto acids in solution NMR sample preps, the Meier group produced methyl proton labeled ubiquitin and demonstrated a 4D HCCH pulse sequence for determining long-range distances in a protein.[69] These restraints where used to solve the structure of ubiquitin (PDB ID: 2L3Z). In addition to nanocrystalline systems, ^{1}H-detection has now been applied in some initial examples to membrane proteins and fibrils.[70]

1.6 Additional New Methods

Another method for long-range NMR distances relies on the paramagnetic relaxation effect (PRE), in which the effect of an unpaired electron spin on the R_1 and R_{1rho} relaxation rates of nearby spins is observed. As recently demonstrated by the Jaroniec

group, by attaching paramagnetic tags, such as a nitroxide or EDTA chelated Cu, to Cys sites in a protein, the PRE can be observed for many sites throughout a protein.[71] These relaxation effects can be modeled to provide accurate distance information for sites between 10 and 20 Å away from the spin label, which can in turn be used to solve the structure of the protein (Helmus et al., unpublished). The thiol chemistry used to attach paramagnetic tags to unique Cys residues can also be used to attach other moieties, such as CH_3 or CF_3 groups. ^{19}F NMR through such groups or ^{19}F labeled amino acids that can also be used to determine long-range distances.[72, 73]

One continual challenge facing SSNMR is sensitivity. By diluting a protein sample many fold in a biologically relevant bilayer, signals, especially for large proteins, are very weak. Higher field magnet systems and ^{1}H detection can help address this shortcoming, but another method that has matured in recent years is dynamic nuclear polarization (DNP). DNP has been shown to give signal enhancements in model systems up to ~150 times, and in protein systems of 30-60.[74] Such factors would allow factor of 10 dilutions with proteins 3-6 times larger while still maintaining similar signal averaging times. Alternatively, extensive spectral editing through 4D experiments to resolve signals becomes much more feasible with the increased amount of signal. Since DNP takes place at cryogenic temperatures, line broadening is an issue; however, the ability to access more dimensions or target a single signal may help overcome broader lines. For example, the Griffin group was able to resolve signals in the active site of bacteriorhodopsin and investigate photointermediate states.[75]

1.7 Computational methods

While every experimental effort is made to obtain unambiguous long-range distance constraints, in most cases for large proteins, chemical shift degeneracy combined with insufficient sensitivity for 3D and/or 4D spectra, results in many observed correlations being ambiguous. Modern structure determination software has developed algorithms to handle such data. X-PLOR's PASD routine[76] and the CNS addition Solaria[77] both allow for a probability based approach to ambiguous distance restraints. By systematically attempting to assign ambiguous correlations, based on comparisons to intermediate structures, these algorithms are able to determine consensus assignments from large numbers of ambiguous distance restraints.

1.8 Future Directions

As we look towards the future, two big advances are notable. The first is automated chemical shift assignments. Currently many months to years are required to obtain complete site-specific chemical shift assignments for large protein systems. Because these site-specific assignments are necessary for any site-specific distance or structural information (PREs, CSTs, VEANS, etc) they are a necessary first step in structure determination. Work in automating this process has proceeded in the solution NMR community for many years, and is more recently starting to gain momentum within the SSNMR community. [78, 79] Because SSNMR linewidths are generally broader than those in solution NMR, and the likelihood of missing signals is greater, modifications of the algorithms are required. Nevertheless, several programs have recently been shown to work on small nanocrystalline proteins, and work on larger systems is underway. Advances in ^{1}H detected SSNMR should also assist in the adoption of automated

assignments, as the datasets obtained more closely resemble the data obtained in solution NMR.

Finally, the canonical procedure of chemical shift assignments followed by obtaining long-range distances and angles information for the use in simulated annealing structure calculations is being complemented by new methods that work straight from chemical shift information. The programs Rosetta and CS-Rosetta[80] have shown great promise for determining the structures of moderately sized proteins using only solution NMR chemical shifts and/or ambiguous NOE information.[81] In addition, other CS based calculations now show a similar ability to tackle moderately sized systems.[82,83] While these techniques may need further enhancement to work with protein fibrils or lipid associated membrane proteins, refinements have enabled CS-Rosetta to deal with multimeric protein complexes.

1.9 Conclusion

In conclusion, over the past two decades the field of protein structure determination by SSNMR has taken huge leaps forward, moving from small polypeptides to protein complexes, fibrils and membrane proteins. As sensitivity and resolution in SSNMR continue to improve, larger and more complicated systems will continue to become available. The unique ability of SSNMR to access these biologically and medically important systems at atomic resolution will enable SSNMR to become a tool of huge importance to structural biology and medicine. The potential to automate much if not all of the data analysis opens the door for high throughput structure determinations, as we are now seeing in solution NMR and X-ray crystallography.

1.10 References

1. Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H., "Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy." *Nature* **2002,** *420* (6911), 98-102.
2. Lange, A.; Becker, S.; Seidel, K.; Giller, K.; Pongs, O.; Baldus, M., "A concept for rapid protein-structure determination by solid-state NMR spectroscopy." *Angew. Chem. Int. Ed.* **2005,** *44* (14), 2089-2092.
3. Zech, S. G.; Wand, A. J.; McDermott, A. E., "Protein structure determination by high-resolution solid-state NMR spectroscopy: Application to microcrystalline ubiquitin." *J. Am. Chem. Soc.* **2005,** *127*, 8618-8626.
4. Manolikas, T.; Herrmann, T.; Meier, B. H., "Protein structure determination from C-13 spin-diffusion solid-state NMR spectroscopy." *Chimia* **2008,** *62* (5), 439-439.
5. Ferguson, N.; Becker, J.; Tidow, H.; Tremmel, S.; Sharpe, T. D.; Krause, G.; Flinders, J.; Petrovich, M.; Berriman, J.; Oschkinat, H.; Fersht, A. R., "General structural motifs of amyloid protofilaments." *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (44), 16248-53.
6. Iwata, K.; Fujiwara, T.; Matsuki, Y.; Akutsu, H.; Takahashi, S.; Naiki, H.; Goto, Y., "3D structure of amyloid protofilaments of beta2-microglobulin fragment probed by solid-state NMR." *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (48), 18119-24.
7. Franks, W. T.; Wylie, B. J.; Schmidt, H. L. F.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M., "Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (12), 4621-4626.
8. Loquet, A.; Bardiaux, B.; Gardiennet, C.; Blanchet, C.; Baldus, M.; Nilges, M.; Malliavin, T.; Bockmann, A., "3D structure determination of the Crh protein from highly ambiguous solid-state NMR restraints." *J. Am. Chem. Soc.* **2008,** *130* (11), 3579-3589.
9. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H., "Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core." *Science* **2008,** *319* (5869), 1523-1526.
10. Cady, S. D.; Schmidt-Rohr, K.; Wang, J.; Soto, C. S.; DeGrado, W. F.; Hong, M., "Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers." *Nature* **2010,** *463* (7281), 689-U127.

11. Jehle, S.; Rajagopal, P.; Bardiaux, B.; Markovic, S.; Kuhne, R.; Stout, J. R.; Higman, V. A.; Klevit, R. E.; van Rossum, B. J.; Oschkinat, H., "Solid-state NMR and SAXS studies provide a structural basis for the activation of alpha B-crystallin oligomers." *Nat. Struct. Mol. Biol.* **2010,** *17* (9), 1037-U1.
12. Opella, S. J.; Marassi, F. M., "Structure determination of membrane proteins by NMR spectroscopy." *Chem. Rev.* **2004,** *104* (8), 3587-3606.
13. Sinha, N.; Grant, C. V.; Park, S. H.; Brown, J. M.; Opella, S. J., "Triple resonance experiments for aligned sample solid-state NMR of C-13 and N-15 labeled proteins." *J. Magn. Reson.* **2007,** *186* (1), 51-64.
14. Williamson, M. P.; Havel, T. F.; Wüthrich, K., "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H NMR and distance geometry." *J. Mol. Biol.* **1985,** *182*, 295.
15. Wüthrich, K., *NMR of Proteins and Nucleic Acids*. John Wiley and Sons: New York, 1986.
16. Takegoshi, K.; Nakamura, S.; Terao, T., "C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR." *Chem. Phys. Lett.* **2001,** *344* (5-6), 631-637.
17. Szeverenyi, N. M.; Sullivan, M. J.; Maciel, G. E., "Observation of Spin Exchange by Two-Dimensional Fourier-Transform C-13 Cross Polarization-Magic-Angle Spinning." *J. Magn. Reson.* **1982,** *47* (3), 462-475.
18. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., "The Xplor-NIH NMR molecular structure determination package." *J. Magn. Reson.* **2003,** *160* (1), 65-73.
19. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., "Crystallography & NMR system: A new software suite for macromolecular structure determination." *Acta Cryst. D* **1998,** *54*, 905-921.
20. Cornilescu, G.; Delaglio, F.; Bax, A., "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." *J. Biomol. NMR* **1999,** *13* (3), 289-302.
21. LeMaster, D. M.; Kushlan, D. M., "Dynamical mapping of E-coli thioredoxin via C-13 NMR relaxation analysis." *J. Am. Chem. Soc.* **1996,** *118* (39), 9255-9264.
22. Hong, M., "Determination of Multiple phi-Torsion Angles in Proteins by Selective and Extensive 13C Labeling and Two-Dimensional Solid-State NMR." *J. Magn. Reson.* **1999,** *139*, 389-401.

23. Rienstra, C. M.; Hohwy, M.; Mueller, L. J.; Jaroniec, C. P.; Reif, B.; Griffin, R. G., "Determination of multiple torsion-angle constraints in U-^{13}C,^{15}N-labeled peptides: 3D ^{1}H-^{15}N-^{13}C-^{1}H dipolar chemical shift NMR spectroscopy in rotating solids." *J. Am. Chem. Soc.* **2002,** *124* (40), 11908-11922.
24. Nieuwkoop, A. J.; Wylie, B. J.; Franks, W. T.; Shah, G. J.; Rienstra, C. M., "Atomic resolution protein structure determination by three-dimensional transferred echo double resonance solid-state nuclear magnetic resonance spectroscopy." *J. Chem. Phys.* **2009,** *131* (9), 095101.
25. Wylie, B. J.; Schwieters, C. D.; Oldfield, E.; Rienstra, C. M., "Protein Structure Refinement Using 13Cα Chemical Shift Tensors." *J. Am. Chem. Soc.* **2009,** *131* (3), 985-992.
26. Wylie, B. J.; Sperling, L. J.; Nieuwkoop, A. J.; Franks, W. T.; Oldfield, E.; Rienstra, C. M., "Ultra-high resolution protein structures using NMR chemical shift tensors." **2011,** *In Press*.
27. Tycko, R., "Solid-state NMR as a probe of amyloid fibril structure." *Curr. Opin. Struct. Biol.* **2000,** *4* (5), 500-506.
28. Tycko, R., Solid-state nuclear magnetic resonance techniques for structural studies of amyloid fibrils. In *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B*, Academic Press Inc: San Diego, 2001; Vol. 339, pp 390-413.
29. Tycko, R.; Petkova, A.; Oyler, N.; Chan, C. C.; Balbach, J., "Probing the molecular structure of amyloid fibrils with solid state NMR." *Biophys. J.* **2002,** *82* (1), 187A-187A.
30. Tycko, R., "Progress towards a molecular-level structural understanding of amyloid fibrils." *Curr. Opin. Struct. Biol.* **2004,** *14* (1), 96-103.
31. Tycko, R., "Molecular structure of amyloid fibrils: insights from solid-state NMR." *Q. Rev. Biophys.* **2006,** *39* (1), 1-55.
32. Tycko, R., "Solid-state NMR as a probe of amyloid structure." *Protein Pept. Lett.* **2006,** *13* (3), 229-234.
33. Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R., "A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR." *Proc. Natl. Acad. Sci. USA* **2002,** *99* (26), 16742-16747.
34. Petkova, A. T.; Leapman, R. D.; Guo, Z. H.; Yau, W. M.; Mattson, M. P.; Tycko, R., "Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils." *Science* **2005,** *307* (5707), 262-265.
35. Paravastu, A. K.; Qahwash, I.; Leapman, R. D.; Meredith, S. C.; Tycko, R., "Seeded growth of beta-amyloid fibrils from Alzheimer's brain-derived fibrils

produces a distinct fibril structure." *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (18), 7443-7448.
36. Jaroniec, C. P.; MacPhee, C. E.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G., "High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy." *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (3), 711-716.
37. Griffin, R. G.; Caporini, M. A.; Bajaj, V. S.; Veshtort, M.; Fitzpatrick, A.; MacPhee, C. E.; Vendruscolo, M.; Dobson, C. M., "Accurate Determination of Interstrand Distances and Alignment in Amyloid Fibrils by Magic Angle Spinning NMR." *J. Phys. Chem. B* **2010,** *114* (42), 13555-13561.
38. Van Melckebeke, H.; Wasmer, C.; Lange, A.; Ab, E.; Loquet, A.; Bockmann, A.; Meier, B. H., "Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state NMR spectroscopy." *J. Am. Chem. Soc.* **2010,** *132* (39), 13765-75.
39. Lewandowski, J. R.; De Paepe, G.; Griffin, R. G., "Proton assisted insensitive nuclei cross polarization." *J. Am. Chem. Soc.* **2007,** *129* (4), 728-9.
40. Lange, A.; Luca, S.; Baldus, M., "Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids." *J. Am. Chem. Soc.* **2002,** *124*, 9704-9705.
41. Hing, A.; Vega, S.; Schaefer, J., "Transferred-echo double-resonance NMR." *J. Magn. Reson.* **1992,** *96*, 205-209.
42. Jaroniec, C. P.; Filip, C.; Griffin, R. G., "3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon-nitrogen distances in uniformly C-13, N-15-labeled solids." *J. Am. Chem. Soc.* **2002,** *124* (36), 10728-10742.
43. Yang, J.; Tasayco, M. L.; Polenova, T., "Magic angle spinning NMR experiments for structural studies of differentially enriched protein interfaces and protein assemblies." *J. Am. Chem. Soc.* **2008,** *130* (17), 5798-5807.
44. Nieuwkoop, A. J.; Rienstra, C. M., "Supramolecular Protein Structure Determination by Site-Specific Long-Range Intermolecular Solid State NMR Spectroscopy." *J. Am. Chem. Soc.* **2010,** *132* (22), 7570-+.
45. Caffrey, M., "Membrane protein crystallization." *J. Struct. Biol.* **2003,** *142* (1), 108-32.
46. Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C., "High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor." *Science* **2007,** *318* (5854), 1258-1265.

47. Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Yao, X. J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K., "GPCR engineering yields high-resolution structural insights into beta(2)-adrenergic receptor function." *Science* **2007,** *318* (5854), 1266-1273.
48. Himber, J.; Refino, C. J.; Burcklen, L.; Roux, S.; Kirchhofer, D., "Inhibition of arterial thrombosis by a soluble tissue factor mutant and active site-blocked factors IXa and Xa in the guinea pig." *Thromb. Haemost.* **2001,** *85* (3), 475-481.
49. Shi, L. C.; Ahmed, M. A. M.; Zhang, W. R.; Whited, G.; Brown, L. S.; Ladizhansky, V., "Three-Dimensional Solid-State NMR Study of a Seven-Helical Integral Membrane Proton Pump-Structural Insights." *J. Mol. Biol.* **2009,** *386* (4), 1078-1093.
50. Kamihira, M.; Vosegaard, T.; Mason, A. J.; Straus, S. K.; Nielsen, N. C.; Watts, A., "Structural and orientational constraints of bacteriorhodopsin in purple membranes determined by oriented-sample solid-state NMR spectroscopy." *J. Struct. Biol.* **2005,** *149* (1), 7-16.
51. Higman, V. A.; Varga, K.; Aslimovska, L.; Judge, P. J.; Sperling, L. J.; Rienstra, C. M.; Watts, A., "The Conformation of Bacteriorhodopsin Loops in Purple Membranes resolved by Solid-state MAS-NMR. ." *Angew. Chem. Int. Ed.* **2011,** *Accepted*.
52. Li, Y.; Berthold, D. A.; Gennis, R. B.; Rienstra, C. M., "Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy." *Protein Sci.* **2008,** *17* (2), 199-204.
53. Tang, M.; Berthold, D. A.; Rienstra, C. M., "Solid-State NMR of a Large Membrane Protein by Paramagnetic Relaxation Enhancement." *J. Phys. Chem. Lett.* **2011,** *2*, 1836-1841.
54. Mani, R.; Tang, M.; Wu, X.; Buffy, J. J.; Waring, A. J.; Sherman, M. A.; Hong, M., "Membrane-Bound Dimer Structure of a beta-Hairpin Antimicrobial Peptide from Rotational-Echo Double-Resonance Solid-State NMR." *Biochemistry* **2006,** *45* (27), 8341-8349.
55. Tang, M.; Hong, M., "Structure and mechanism of beta-hairpin antimicrobial peptides in lipid bilayers from solid-state NMR spectroscopy." *Mol. BioSyst.* **2009,** *5* (4), 317-22.
56. Tang, M.; Sperling, L. J.; Berthold, D. A.; Schwieters, C. D.; Nesbitt, A. E.; Nieuwkoop, A. J.; Gennis, R. B.; Rienstra, C. M., "High-resolution membrane protein structure by joint calculations with solid-state NMR and X-ray experimental data." **2011,** *Submitted*.

57. Frericks, H. L.; Zhou, D. H.; Yap, L. L.; Gennis, R. B.; Rienstra, C. M., "Magic-angle spinning solid-state NMR of a 144 kDa membrane protein complex: E-coli cytochrome bo(3) oxidase." *J. Biomol. NMR* **2006,** *36* (1), 55-71.
58. Bodenhausen, G.; Ruben, D. J., "Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy." *Chem. Phys. Lett.* **1980,** *69* (1), 185-189.
59. Paulson, E. K.; Morcombe, C. R.; Gaponenko, V.; Dancheck, B.; Byrd, R. A.; Zilm, K. W., "Sensitive high resolution inverse detection NMR spectroscopy of proteins in the solid state." *J. Am. Chem. Soc.* **2003,** *125* (51), 15831-15836.
60. Morcombe, C. R.; Paulson, E. K.; Gaponenko, V.; Byrd, R. A.; Zilm, K. W., "H-1-N-15 correlation spectroscopy of nanocrystalline proteins." *J. Biomol. NMR* **2005,** *31* (3), 217-230.
61. Reif, B.; Griffin, R. G., "H-1 detected H-1, N-15 correlation spectroscopy in rotating solids." *J. Magn. Reson.* **2003,** *160* (1), 78-83.
62. Agarwal, V.; Diehl, A.; Skrynnikov, N.; Reif, B., "High Resolution 1H Detected 1H,13C Correlation Spectra in MAS Solid-State NMR using Deuterated Proteins with Selective 1H,2H Isotopic Labeling of Methyl Groups." *J. Am. Chem. Soc.* **2006,** *128* (39), 12620-12621.
63. Zhou, D. H.; Shah, G.; Cormos, M.; Mullen, C.; Sandoz, D.; Rienstra, C. M., "Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning." *J. Am. Chem. Soc.* **2007,** *129* (38), 11791-11801.
64. Zhou, D. H.; Rienstra, C. M., "High-performance solvent suppression for proton detected solid-state NMR." *J. Magn. Reson.* **2008,** *192* (1), 167-172.
65. Chevelkov, V.; Rehbein, K.; Diehl, A.; Reif, B., "Ultra-High Resolution in Proton Solid-State NMR Spectroscopy at High Levels of Deuteration." *Angew. Chem. Int. Ed.* **2006,** *45* (23), 3878-3881.
66. Akbey, U.; Lange, S.; Franks, W. T.; Linser, R.; Rehbein, K.; Diehl, A.; van Rossum, B. J.; Reif, B.; Oschkinat, H., "Optimum levels of exchangeable protons in perdeuterated proteins for proton detection in MAS solid-state NMR spectroscopy." *J. Biomol. NMR* **2010,** *46* (1), 67-73.
67. Zhou, D. H.; Shea, J. J.; Nieuwkoop, A. J.; Franks, W. T.; Wylie, B. J.; Mullen, C.; Sandoz, D.; Rienstra, C. M., "Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy." *Angew. Chem., Int. Ed.* **2007,** *46* (44), 8380-8383.
68. Tugarinov, V.; Kay, L. E., "Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods." *J. Am. Chem. Soc.* **2003,** *125* (45), 13868-13878.

69. Huber, M.; Hiller, S.; Schanda, P.; Ernst, M.; Bockmann, A.; Verel, R.; Meier, B. H., "A proton-detected 4D solid-state NMR experiment for protein structure determination." *ChemPhysChem* **2011,** *12* (5), 915-8.
70. Linser, R.; Dasari, M.; Hiller, M.; Higman, V.; Fink, U.; del Amo, J. M. L.; Markovic, S.; Handel, L.; Kessler, B.; Schmieder, P.; Oesterhelt, D.; Oschkinat, H.; Reif, B., "Proton-Detected Solid-State NMR Spectroscopy of Fibrillar and Membrane Proteins." *Angew. Chem., Int. Ed.* **2011,** *50* (19), 4508-4512.
71. Nadaud, P. S.; Helmus, J. J.; Kall, S. L.; Jaroniec, C. P., "Paramagnetic Ions Enable Tuning of Nuclear Relaxation Rates and Provide Long-Range Structural Restraints in Solid-State NMR of Proteins." *J. Am. Chem. Soc.* **2009,** *131* (23), 8108-8120.
72. McDowell, L. M.; Lee, M. S.; McKay, R. A.; Anderson, K. S.; Schaefer, J., "Intersubunit communication in tryptophan synthase by carbon-13 and flourine-19 REDOR NMR." *Biochemistry* **1996,** *35* (10), 3328-3334.
73. Graesser, D. T.; Wylie, B. J.; Nieuwkoop, A. J.; Franks, W. T.; Rienstra, C. M., "Long-range ^{19}F-^{15}N distance measurements in highly-^{13}C,^{15}N-enriched solid proteins with ^{19}F-dephased REDOR shift (FRESH) spectroscopy." *Magn. Reson. Chem.* **2007,** *45*, S129-S134.
74. Song, C.; Hu, K. N.; Joo, C. G.; Swager, T. M.; Griffin, R. G., "TOTAPOL: a biradical polarizing agent for dynamic nuclear polarization experiments in aqueous media." *J. Am. Chem. Soc.* **2006,** *128* (35), 11385-90.
75. Mak-Jurkauskas, M. L.; Bajaj, V. S.; Hornstein, M. K.; Belenky, M.; Griffin, R. G.; Herzfeld, J., "Energy transformations early in the bacteriorhodopsin photocycle revealed by DNP-enhanced solid-state NMR." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (3), 883-8.
76. Kuszewski, J.; Schwieters, C. D.; Garrett, D. S.; Byrd, R. A.; Tjandra, N.; Clore, G. M., "Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments." *J. Am. Chem. Soc.* **2004,** *126* (20), 6258-6273.
77. Fossi, M.; Castellani, F.; Nilges, M.; Oschkinat, H.; van Rossum, B. J., "SOLARIA: a protocol for automated cross-peak assignment and structure calculation for solid-state magic-angle spinning NMR spectroscopy." *Angew. Chem. Int. Ed.* **2005,** *44* (38), 6151-4.
78. Tycko, R.; Hu, K. N., "A Monte Carlo/simulated annealing algorithm for sequential resonance assignment in solid state NMR of uniformly labeled proteins with magic-angle spinning." *J. Magn. Reson.* **2010,** *205* (2), 304-314.

79. Moseley, H. N. B.; Sperling, L. J.; Rienstra, C. M., "Automated protein resonance assignments of magic angle spinning solid-state NMR spectra of beta 1 immunoglobulin binding domain of protein G (GB1)." *J. Biomol. NMR* **2010,** *48* (3), 123-128.
80. Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A., "Consistent blind protein structure generation from NMR chemical shift data." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (12), 4685-4690.
81. Raman, S.; Huang, Y. J. P.; Mao, B. C.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Montelione, G. T.; Baker, D., "Accurate Automated Protein NMR Structure Determination Using Unassigned NOESY Data." *J. Am. Chem. Soc.* **2010,** *132* (1), 202-207.
82. Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G. H.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D., "NMR Structure Determination for Larger Proteins Using Backbone-Only Data." *Science* **2010,** *327* (5968), 1014-1018.
83. Shen, Y.; Bryan, P. N.; He, Y. N.; Orban, J.; Baker, D.; Bax, A., "De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds." *Protein Sci.* **2010,** *19* (2), 349-356.

Chapter 2

Atomic Resolution Protein Structure Determination by 3D TEDOR Solid-State NMR Spectroscopy

2.1 Notes and Acknowledgements

This chapter is reprinted with permission from “Atomic resolution protein structure determination by three- dimensional transferred echo double resonance solid-state nuclear magnetic resonance spectroscopy", Nieuwkoop, A. J.; Wylie, B. J.; Franks, W. T.; Shah, G. J.; Rienstra C. M. *J. Chem. Phys.*, **2009**, *131*, 095101. Copyright 2009, American Institute of Physics.

This work was supported by the National Institutes of Health R01-GM073770. Andy Nieuwkoop was supported by the Molecular Biophysics Training Program at the University of Illinois.

2.2 Introduction

Magic-angle spinning (MAS) solid-state NMR (SSNMR) has been used to solve several complete protein structures in recent years,[1-6] and the precision of these structures is improving.[5] These advances are important because SSNMR is able to investigate samples with only local order and offers the most promising avenue towards routinely solving the structures of insoluble proteins that do not form single crystals. Specifically, substantial progress has been made in SSNMR structural investigations of both membrane proteins[7-14] and protein aggregates.[6, 15-20] One major remaining goal is to solve structures of these systems at a rate and quality comparable to those produced by solution NMR and x-ray crystallography. Most SSNMR structures so far have been solved with a combination of semi-quantitative ^{13}C-^{13}C, ^{15}N-^{15}N and ^{1}H-^{1}H distances, of precision comparable to nuclear Overhauser effects (NOEs) measured in solution. In addition, empirical dihedral angle restraints from TALOS assist in computational refinement of secondary structure elements.[21] Until recently, the highest quality SSNMR structures were reported with backbone RMSD (bbRMSD) values of ~0.8 to 1.4 Å.[1-4] We have recently reported that with large numbers of distances, as well as high precision vector angle restraints, structures can be refined to ~0.2 Å bbRMSD.[5] Likewise, we have shown that chemical shift tensors provide another avenue to atomic resolution structure refinement.[22]

Here we investigate the measurement of high precision heteronuclear (^{15}N-^{13}C) distance restraints to refine protein structure to atomic-resolution. Heteronuclear dipolar couplings, normally averaged by MAS, can be reintroduced with a train of rotor synchronized rf-pulses, as in the rotational echo double resonance (REDOR)[23]

experiment and its analog transferred echo double resonance (TEDOR).[24] For spin pairs or clusters, these methods offer exquisitely high precision determination of internuclear distances by reporting dipolar dephasing (for REDOR) or buildup (for TEDOR) trajectories under the influence of a pulse sequence in which the multiple spin-pair interactions mutually commute with one another.[25, 26] Furthermore, for application to proteins with a large number of isotopic labels, the 3D z-filtered TEDOR (ZF-TEDOR)[27] pulse sequence was developed, enabling quantitative trajectories to be extracted from a series of 2D spectra in which the ^{13}C and ^{15}N chemical shifts for each unique spin pair are encoded. Therefore, many dozens of distances could be measured in uniformly-^{13}C,^{15}N-labeled peptides.[28, 29]

Although the ZF-TEDOR experiment addresses the direct and indirect effects of ^{13}C-^{13}C scalar couplings in the observed trajectories and peak lineshapes, the modulations arising from one-bond couplings compromise the dynamic range of the measurement for weak dipolar couplings. This well-recognized challenge has previously been addressed by band-selective decoupling (BASE TEDOR),[27] as well as a semi-constant time TEDOR version.[30] Here we address this in a third, complementary way, by preparing protein samples where the percentage of directly bonded ^{13}C pairs is minimized, as derived from either 1,3-^{13}C-glycerol (along with natural abundance carbonate) or 2-^{13}C-glycerol (and ^{13}C carbonate) as its sole sources of carbon. This expression scheme, originally developed by LeMaster *et al.*,[31] and utilized previously in SSNMR,[1, 5, 32] produces an anticorrelated "checkerboard" pattern of ^{13}C and ^{12}C labeling for most amino acids, enhancing resolution and sensitivity most notably in ^{13}C-^{13}C 2D and ^{15}N-^{13}C-^{13}C 3D experiments.

We show here that for 3D ZF-TEDOR experiments, this glycerol-derived labeling pattern is particularly beneficial, enabling high-quality experimental data to be acquired with hundreds of resolved correlations. We then demonstrate a protocol for fitting the dipolar trajectories quantitatively, modeling the spin dynamics for each set of several ^{15}N spins coupled to each resolved ^{13}C site. This approach enables quantitative analysis for distances of up to at least ~5 Å, as well as detection of much longer distances (~8 Å) with moderate precision. These distance restraints prove useful in determining protein tertiary structure.

2.3 Experimental and Computational Methods

2.3.1 Sample Preparation.

Samples of GB1, a 6 kDa *streptococcil protein expressed from E. coli* were ^{13}C and ^{15}N isotopically labeled by bacterial overexpression in media containing ^{15}N ammonium chloride as the sole nitrogen source and either (a) 2-^{13}C-glycerol and calcium ^{13}C-carbonate, or (b) 1,3-^{13}C-glycerol and natural abundance carbonate, as the sole carbon sources.[1, 3, 5, 31, 33] We refer to these preparations throughout the text as [2]-GB1 and [1,3]-GB1 respectively. In each case, ~18 mg of nanocrystalline protein was packed into the central 80% of a limited speed 3.2 mm Varian rotor (Varian, Inc., Fort Collins, Colorado). All experiments were performed using a 500 MHz InfinityPlus spectrometer (Varian, Inc., Palo Alto, California and Fort Collins, Colorado) equipped with a 3.2 mm T3 Balun^{TM} ^{1}H-^{13}C-^{15}N MAS probe. Pulse widths (π/2) for ^{1}H, ^{13}C, and ^{15}N were 2.6 ms, 3.0 ms, and 6.0 ms respectively. Spinning was controlled via a Varian MAS controller to 11,111 ± 2 Hz.

2.3.2 NMR Spectroscopy.

A series of two-dimensional (2D) ^{15}N-^{13}C planes was acquired according to the 3D ZF-TEDOR pulse sequence.[27] TPPM decoupling was used during acquisition with a ^{1}H field of ~70 kHz (6.7 ms, 15.0^{o} total phase difference). TPPM conditions during REDOR (5.0 ms, 18.0^{o}) were optimized with 5.76 ms mixing and a nominal ^{1}H field of ~100 kHz. ^{13}C and ^{15}N p-pulse widths during TEDOR were 6.0 ms and 15.2 ms respectively, the latter value adjusted to ensure that the ratio of ^{1}H to ^{15}N nutation frequency was at least 3:1 to minimize decoupling interference.[34, 35] Spectra were acquired with the minimum phase cycle of 16 scans per row, resulting in 4.5 hour blocks of measurement time with each 2D plane digitized to 640 points (TPPI) by 45 μs per row in t_1 (^{15}N, 28.8 ms maximum evolution time) and 3072 complex points (15 μs dwell) in the t_2 acquisition dimension (^{13}C, 46.1 ms acquisition time). Data processing was performed with 15 Hz and 5 Hz net line broadening in the direct and indirect dimensions respectively (using a 1:2.5 ratio of negative Lorentzian to positive Gaussian apodization) and zero filling to 8192 by 8192 complex points.

The ^{15}N-^{13}C dipolar trajectories were sampled with values of t_{mix} (according to Jaroniec et al., Figure 2.1)[27] incremented from 1.44 ms to 14.40 ms in steps of 1.44 ms. Signal averaging times were increased for longer mixing times, with a minimum of two blocks (9 h) acquired with t_{mix} = 1.44 ms and a maximum of nine blocks (40.5 h) with t_{mix} = 14.40 ms. Spectra were normalized, according to the total number of scans per row, prior to further analysis.

2.3.3 Numerical Simulations and Data Fitting.

Spectra were processed in NMRPipe[36] and peak intensities extracted from the 2D ^{15}N-^{13}C planes using the nlinLS package, resulting in trajectories of integrated peak intensities as a function of mixing time, from 1.44 ms to 14.4 ms in 1.44 ms increments.

Simulated trajectories were generated using SPINEVOLUTION[37] to model spin dynamics of ZF-TEDOR with the following assumptions: (1) All spin coherences derived from a single ^{13}C site commute with all other spin operators involving other ^{13}C spins; (2) each ^{13}C site is coupled to *n* ^{15}N sites, which are not coupled to other ^{15}N spins; (3) the relative orientations of the heteronuclear dipoles could be ignored. (We show experimentally (*vide infra*) that these approximations are valid in the limit of weak couplings.) Thus, simulated trajectories for each of the *n* ^{15}N sites coupled to a single ^{13}C resonance were generated.

The dipolar couplings and other incidental parameters were then determined by minimizing the global difference between the simulated and experimental trajectories, using in-house FORTRAN code that called MINUIT minimization libraries and SPINEVOLUTION.[37] The distance between ^{13}C and ^{15}N spin pairs was simulated for each point on the trajectory and a scaling factor was applied to the result to account for the ^{13}C labeling percentage. In this scheme, y-scaling, corresponding to labeling percentage, is held fixed over the ensemble and T_2 relaxation, applied as a exponential decay, is allowed to vary by 10% over the ensemble, to account for possible higher order relaxation from ^{15}N CSA recoupling during the REDOR periods.

2.4 2D TEDOR Heteronuclear Correlation Spectroscopy

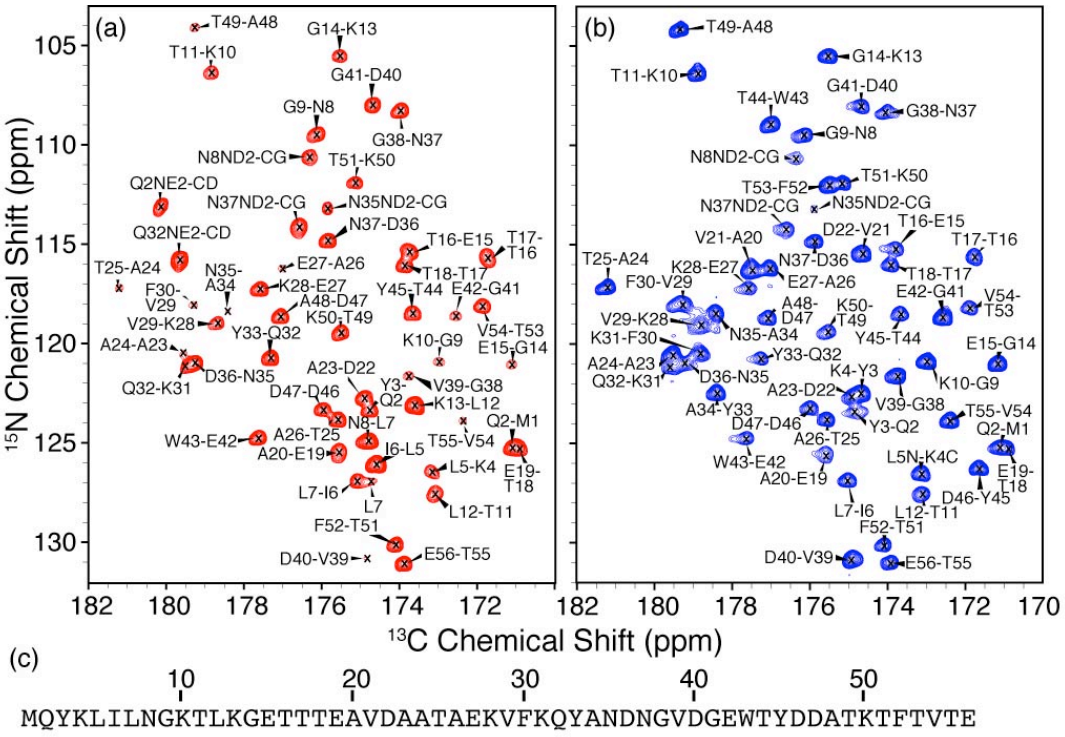

Figure 2.1: Carbonyl ^{13}C region of 2D ^{15}N-^{13}C TEDOR spectra

Spectra of GB1 samples prepared from (a) 2-^{13}C-glycerol, calcium ^{13}C-carbonate and $^{15}NH_4Cl$ ([2]-GB1) and (b) 1,3-^{13}C-glycerol, natural abundance carbonate and $^{15}NH_4Cl$ ([1,3]-GB1). The sequence of GB1 (c) is provided for reference. Spectra were acquired at 500 MHz ^{1}H frequency with 1.44 ms of TEDOR mixing. Peaks are labeled with their backbone ^{15}N and ^{13}C' frequencies respectively, except for sidechain ^{15}N sites which are explicitly indicated. Data were processed with 15 Hz net line broadening (Lorentzian-to-Gaussian apodization) in ^{13}C and 5 Hz in ^{15}N. Acquisition time was 9 hours for each spectrum.

As demonstrated previously,[38] the 2D ^{15}N-^{13}C' spectra of [2]-GB1 (Figure 2.1a) and [1,3]-GB1 (Figure 2.1b) GB1 are well resolved, with natural ^{13}C linewidths of ~0.2 ppm and ^{15}N linewidths of ~0.5 ppm, principally limited by instrumental factors (such as the B_0 homogeneity and the maximum time for which high power ^{1}H decoupling can be applied). Spectra acquired at 1.44 ms ^{15}N-^{13}C mixing, the shortest value utilized here, contain strong peaks from one-bond correlations (e.g., N[i]-C'[i-1]), with the relative intensity depending principally upon the efficiency of the ^{13}C labeling within each amino

acid. For example, in the [2]-GB1 sample, carbonyl sites for Leu are nearly 100% labeled, whereas Ala, Gly, Val and other amino acids from the glycolysis pathway are labeled only to a small extent; in the [1,3]-GB1 sample, a complementary pattern is observed. In both samples, carbonyl sites from amino acids in the citric acid cycle (most notably, Gln/Glu, Asp/Asn, Lys and Thr) are observed with approximately equal intensity in the [2]-GB1 and [1,3]-GB1 samples, due to the scrambling of ^{13}C within the cycle. These signals serve as valuable internal controls to validate data analysis procedures.

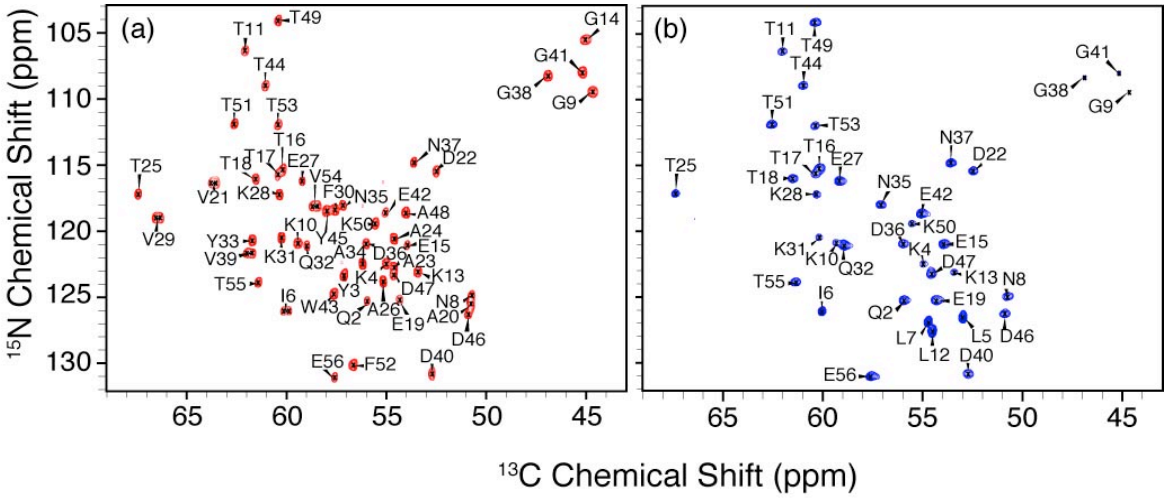

Figure 2.2: CA region of 2D ^{15}N-^{13}C TEDOR spectra.

(a) [2]-GB1; (b) [1,3]-GB1. Spectra acquired with 1.44 ms of TEDOR mixing and were processed with 15 Hz net line broadening (Lorentzian-to-Gaussian apodization) in ^{13}C and 5 Hz in ^{15}N. Acquisition time was 9 hours for each spectrum.

Likewise, the TEDOR spectra are well resolved in the CA region, for both the [2]-GB1 (Figure 2.2a) and [1,3]-GB1 (Figure 2.2b) samples, with nearly every expected cross peak resolved. Partial overlap of the N8, A20 and D46 resonances in the [2]-GB1 spectrum is alleviated by the reduction of A20 intensity in the [1,3]-GB1 spectrum. An additional benefit of the glycerol labeling scheme is the absence of most one-bond homonuclear J-couplings, with the exception the Val and Ile CA-CB, which lead to

doublet patterns observed for I6 and the four Val residues (21, 29, 39 and 54). Moreover, broadening from off or near rotational resonance (R^2) is avoided.[33, 39-43] These factors together contribute to high observed signal-to-noise ratio (SNR) for both samples, approximately 500:1 in nine hours of data collection for directly bonded ^{15}N-^{13}C pairs that were ~100% labeled, and proportionately lower for the fractionally labeled sites. Thus it was also possible to observe weak peaks at natural abundance and/or with a small percentage of labeling; in several instances, sites that are nominally unlabeled appeared with approximately 5% of the maximum intensity (e.g., A48 C' in the [2]-GB1 sample, Figure 2.1a).

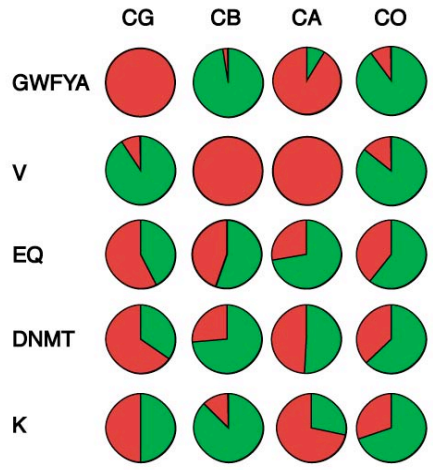

Figure 2.3: ^{13}C labeling pattern measured by TEDOR in GB1.

Red represents the [2-GB1] and green [1,3-GB1]. Each circle represents the intensity of correlations to the carbon type indicated at the top in the residue group indicated to the left.

The glycerol-derived isotopic incorporation on an amino acid specific basis is likely to vary among proteins, due to differences in the relative percentages of each amino acids; in the case of GB1, such effects are likely to be exaggerated by the very high levels of expression (>100 mg / L) and the fact that GB1 lacks several amino acids (Cys, His, Pro, Ser, Arg) and has a large number of others (such as Thr, Asp and Glu).

Therefore the NCA and NCO 2D spectra can be utilized to approximate the percentage of ^{13}C labeling in our samples. This percentage was incorporated as a parameter into the fitting procedure as an intensity scaling factor; this treatment is valid in the limit that the relaxation parameters are similar for all sites, which is satisfactory in the limit of short TEDOR mixing times. The results of this analysis are shown in Figure 2.3. Overall patterns were very similar to those reported previously,[1, 3] but with substantial signal intensities (~10% of the maximum) for sites that were not expected to be labeled, such as CO resonances for Gly, Trp, Phe, Tyr and Ala in the [2]-GB1 sample. Despite the presence of these peaks, we consistently observed lineshapes lacking fine structure from $^{1}J_{CC}$ couplings, consistent with the anticorrelated labeling of neighboring carbons.

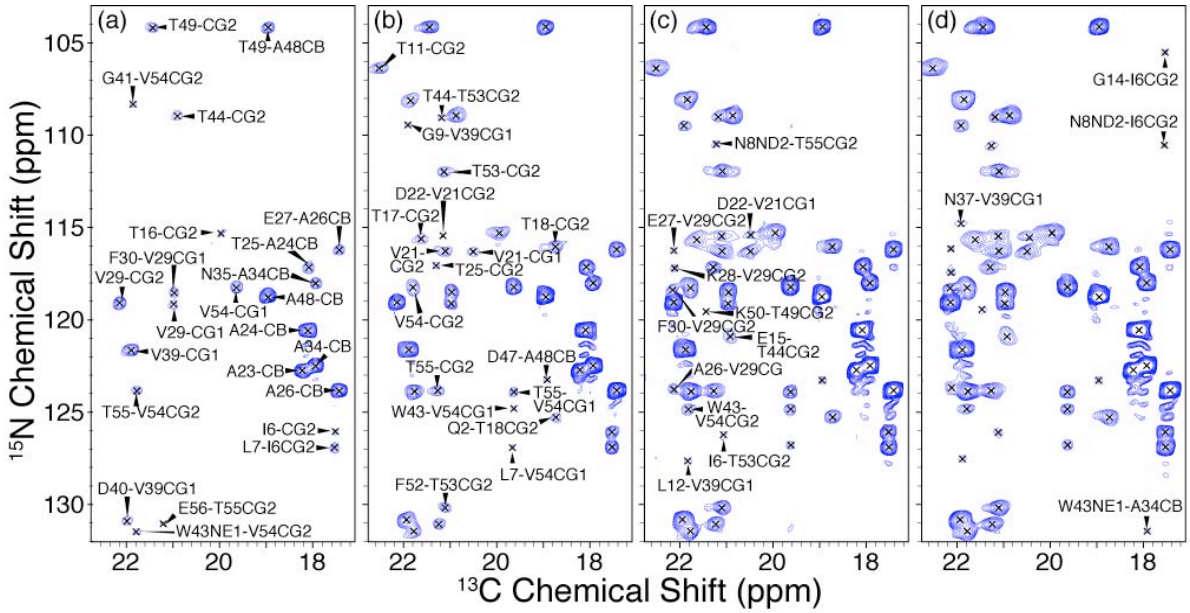

Figure 2.4: Methyl region of 2D ^{15}N-^{13}C TEDOR spectrum of [1,3]-GB1 at several mixing times.

(a) 1.44 ms; (b) 2.88 ms; (c) 4.32 ms; (d) 5.76 ms. Peaks are labeled only in the spectrum where they first appear. Labels indicate the backbone ^{15}N site followed by the ^{13}C to which it is correlated, except for side chain ^{15}N sites that are specifically noted. Intraresidue correlations have no second residue number before the ^{13}C label. Spectra were processed with 15 Hz net line broadening (Lorentzian-to-Gaussian apodization) in ^{13}C and 5 Hz in ^{15}N.

At longer ^{15}N-^{13}C mixing times, additional ^{13}C resonances are observed at each ^{15}N frequency (Figure 2.4); many of these correlations are consistent with medium and long-range distance restraints. For example, the methyl region of [1,3]-GB1 contains strong crosspeaks for two-bond N[i]-CB[i] pairs (dipolar coupling ~200 Hz) with SNR values of ~30, or ~6% of the maximum intensity for the aforementioned one-bond correlations at 1.44 ms ^{15}N-^{13}C mixing. At progressively longer mixing times, these crosspeaks increase in intensity to a maximum of ~150 at 5.76 ms (~30% of the maximum one-bond crosspeak intensity). Even at the 1.44 ms mixing time, some interresidue correlations are observed (e.g., D40-V39CG1, G41-V54CG2, W43NE1-V54CG2), which

must have distances of greater than 3 Å. Peaks first appearing at 2.88 ms include those reporting on backbone-to-sidechain distances of ~4 Å (Val and Thr N-CG) as well as various long-range correlations (G9-V39CG1, L7-V54CG1, Q2-T18CG2, W43-V54CG1). Nearly every ^{13}C and ^{15}N frequency is uniquely resolved, enabling straightforward assignment of the longer mixing time spectra. Chemical shifts agreed well with previously published values.[44]

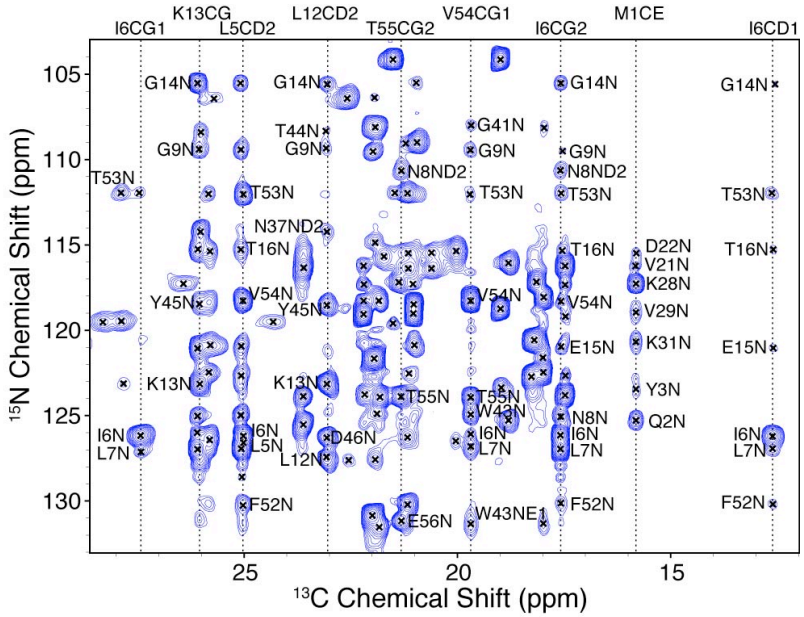

Figure 2.5. Methyl region of [1,3]-GB1 2D ^{15}N-^{13}C TEDOR spectrum.

Spectrum acquired with 14.4 ms ^{15}N-^{13}C mixing, 40 hours measurement time. Data were processed with 15 Hz net line broadening (Lorentzian-to-Gaussian apodization) in ^{13}C and 5 Hz in ^{15}N.

To evaluate the distance measurement range using this approach, we examined the spectrum at 14.4 ms mixing, the longest utilized in this study. At selected ^{13}C frequencies, correlations are observed to nearly a dozen ^{15}N sites. For example, I6CD1 (a uniquely resolved ^{13}C frequency, Figure 2.5) shows correlations not only within the b1 strand to L5, I6 and L7, but also to b2 (G14, E15, and T16) and b4 (F52 and T53).

These correlations are potentially useful in restraining the relative position and register of these three secondary structure elements. Likewise, L5CD2 exhibits correlations within the b1 strand (to L5/L7, I6), b2 (T16) and b4 (F52 and T53). Complementary information is observed for V54CG1, within b4, with correlations to its neighboring residues (T53 and T55), b1 (I6, L7, G9), and b3 (G41, W43). Such correlations between secondary structure elements are reliably observed, and in several cases unambiguous peaks corresponding to distances of 8 Å or more (based on the 2QMT x-ray crystal structure[45]) are observed. Examples shown in Figure 2.5 include F52N-I6CG1 (7.79 Å distance in the crystal structure), V21N-M1CE (8.43 Å) and G14N-I6CD1 (8.71 Å). These long-range peaks are the most useful with regards to defining the tertiary structure of the protein.

We considered whether some of the observed cross peaks might arise due to a multi-step polarization transfer involving two ^{13}C spins and one ^{15}N. This seemed unlikely due to the fact that the MAS rate was chosen to avoid R^2 conditions.[33, 39-43] Furthermore, at the longest mixing time utilized here (i.e., 14.4 ms), with high power decoupling the rate of homonuclear transfer via spin diffusion must be substantially slower than what we observed in GB1 with rotary resonance recoupling on the ^{1}H channel. In that case,[5] we found it necessary to mix at least 50 ms to observe significant (>5% of the diagonal) intensity between ^{13}C pairs separated by 2.5 Å. We examined a specific case, of Ile6 CD1, in the resulting GB1 structure to determine whether there were additional ^{13}C sites physically positioned between source (Ile6 CD1) and destination (G14N) spins; in this case there were none within 5 Å. We attribute this, and

the general lack of spin diffusion by ^{13}C-^{13}C couplings, as a consequence of the sparse, glycerol-derived isotopic labeling.

2.5 Quantitative analysis of polarization transfer trajectories

As illustrated above, a large number of crosspeaks are observed in the 2D TEDOR spectra, and the corresponding distances may be approximated based on relative intensity within individual spectra, in a manner analogous to NOE analysis by solution NMR. However, for purposes of quantitative distance determinations, it is necessary to fit the trajectories with rigorous spin physics models. To do so, we next illustrate the fitting procedure for the polarization transfer trajectories. Like REDOR difference spectra,[23] TEDOR trajectories[24] for spin pairs can readily be fit to analytical (or numerically exact) spin dynamics models to extract accurate ^{15}N-^{13}C distances, and similar logic has been utilized to examine clusters of several spins.[46-48] In contrast to REDOR, however, the fit of the TEDOR trajectories is somewhat complicated by the fact that the maximum absolute intensity in the trajectory depends not only on heteronuclear dipolar coupling, but also the ^{13}C labeling percentage, as well as other factors (relaxation, presence of multiple dipoles, relative orientations, etc.).[27] In particular, the labeling percentage is a critical parameter that contributes to the normalization of the spectral intensity, which distinguishes between (a) short distances involving a fractionally labeled ^{13}C and (b) long distances involving a ^{13}C that is ~100% labeled. These factors must be explicitly considered in order to determine quantitative coupling constants. To first approximation, we considered the couplings, labeling percentage and relaxation as fit parameters, assuming that the labeling percentage and relaxation must be the same for each ^{13}C site, and that all ^{15}N sites were labeled to

100%. The relative orientations of multiple ^{13}C-^{15}N dipoles were ignored.

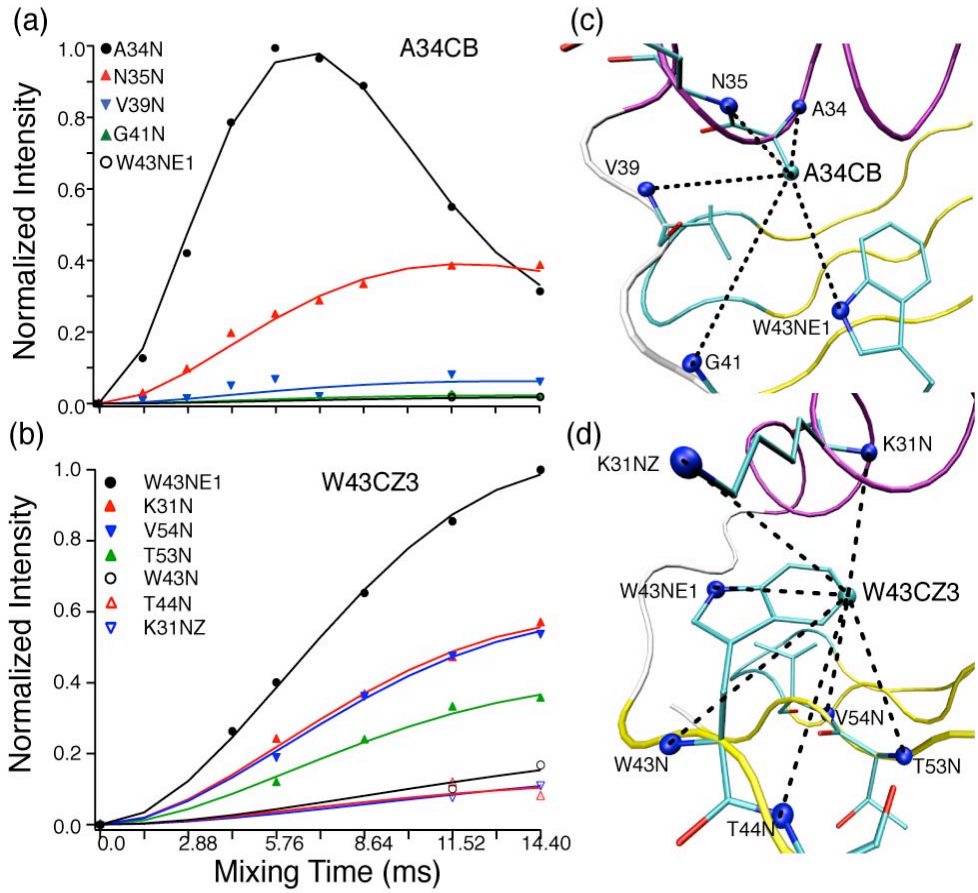

Figure 2.6: TEDOR polarization transfer trajectories selected ^{13}C-^{15}N correlations in GB1.

Trajectories for (a) A34CB in [1,3]-GB1 and (b) W43CZ3 in [2]-GB1. Correlations labeled on the 2QMT crystal structure for (c) A34B and (d) W43CZ3. Intensities in the trajectories are derived from integrated volumes of peaks in 2D ^{15}N-^{13}C TEDOR experiments at the respective mixing times. Simulated trajectories were generated using SPINEVOLUTION simulations of the TEDOR pulse sequence as described in the text. Fit and X-ray distances for A34CB (a,c) are, respectively, A34N, 2.44±0.10 Å v. 2.44 Å; N35N, 3.29±0.10 Å v. 3.09 Å; V39N, 4.46±0.86 Å v. 5.13 Å; G41N, 5.30±1.79 Å v. 6.74 Å; W43NE1, 5.9±3.2 v. 5.47 Å. Fit and X-ray distances for W43CZ3 (b,d) are, respectively, W43NE1. 4.10 v. 4.06 Å; K31N, 4.53 v. 4.61 Å ; V54N, 4.55 v. 4.85 Å; T53N, 4.90 v. 5.18 Å; W43N, 6.02 v. 6.36 Å; T44N, 6.04 v. 6.15 Å; K31NZ, 6.38 v. 6.00 Å. The molecular graphic shows the backbone of helices in purple, strands in yellow and loops in cyan; the dark blue circles are nitrogen atoms, and the distances to the labeled ^{13}C site indicate with dotted lines.

Figure 2.6 illustrates two typical scenarios encountered in fitting ensembles of ^{15}N sites correlated to a single ^{13}C. The A34CB resonance (Figure 2.6a) has intraresidue,

sequential, and long-range correlations. The intraresidue (A34N-A34CB) trajectory exhibits a maximum at 5.76 ms mixing, from which the distance of 2.44 Å is fitted directly, also uniquely constraining the intensity scaling factor. With accurate knowledge of this scaling factor, the remaining trajectories can be fit unambiguously. In comparison to the crystal structure (2QMT, Figure 2.6c), we find reasonable agreement for the short (A34N-A34CB, 2.44±0.10 Å TEDOR v. 2.44 Å X-ray), medium (N35N-A34CB, 3.29±0.10 Å TEDOR v. 3.09 Å X-ray), and long (V39N-A34CB, 4.46±0.86 Å TEDOR v. 5.13 Å X-ray; G41N-A34CB, 5.30±1.79 Å TEDOR v. 6.74 Å X-ray; W43NE1-A34CB, 5.9±3.2 Å TEDOR v. 5.47 Å X-ray) distances. Uncertainties in the TEDOR-determined distances (as determined by the MINUIT routine) are largest for the peaks observed only at the long mixing times, such as W43NE1-A34CB, which is at the borderline of sensitivity in this experiment.

In general, ^{15}N-^{13}C distances of less than ~3.5 Å will yield trajectories exhibit maxima at less than 14.4 ms, which can be uniquely fit to intensity scaling and heteronuclear coupling. In cases where no ^{15}N nuclei are within 3.5 Å of an observed ^{13}C resonance, this procedure becomes problematic and the final scaling factor, and therefore distances, becomes less reliable. For example, the W43CZ3 TEDOR trajectory (Figure 2.6b) has not yet reached its maximum at 14.4 ms; trajectories in the initial rate regime exhibit a strong covariance between the intensity scaling factor and the dipolar coupling. Therefore the certainty of distance determination is limited by knowledge of the hypothetical maximum signal intensity. To address this problem, knowledge of known molecular geometry in this case can be utilized as an internal control to disambiguate the fit parameters. Specifically, the W43NE1-W43CZ3 distance is dictated by the

geometry of the indole ring to be 4.06 Å. Thus in fitting this group of trajectories we constrained this distance to be 4.05 ± 0.05 Å, and the global fit for all six distances was found to be internally consistent. Distances calculated from this trajectory were systematically short without the control, but using this chemical restraint the ^{15}N to W43CZ3 distances converge very well to those observed in the X-ray structure (K31N, 4.53±0.10 v. 4.61 Å X-ray; V54N, 4.55±0.10 v. 4.85 Å; T53N, 4.90±3.90 v. 5.18 Å; W43N, 6.02±0.10 v. 6.36 Å; T44N, 6.04±0.10 v. 6.15 Å; K31NZ, 6.38±0.10 v. 6.00 Å). When fixing a distance within the ensemble, error determinations for the remaining distances become less reliable as the mingrad MINUIT algorithm cannot properly sample the solution space to determine the reliability of the fit, although a reasonable estimate of the uncertainty for all trajectories shown here is better than ±0.4 Å (i.e., the range of distances from the mean required to generate simulations that encompass all observed experimental data points). A total of 726 distance restraints (340 from the [1,3]-GB1 and 386 from the [2]-GB1) were determined. Among these, 454 had distance uncertainties of 0.2 Å or less; another 95 had uncertainties between 0.2 and 0.5 Å, 48 between 0.5 and 1.0 Å, and the remaining 64 had uncertainties greater than 1.0 Å. In some instances (including most of the 64 reported to have greater than 1.0 Å uncertainty), MINUIT failed to report reasonable errors (for example T53N-W43CZ3, Figure 2.6b) although by manual inspection it was clear that the trajectories were well fit. Nevertheless, the large uncertainties were used in subsequent calculations (*vide infra*).

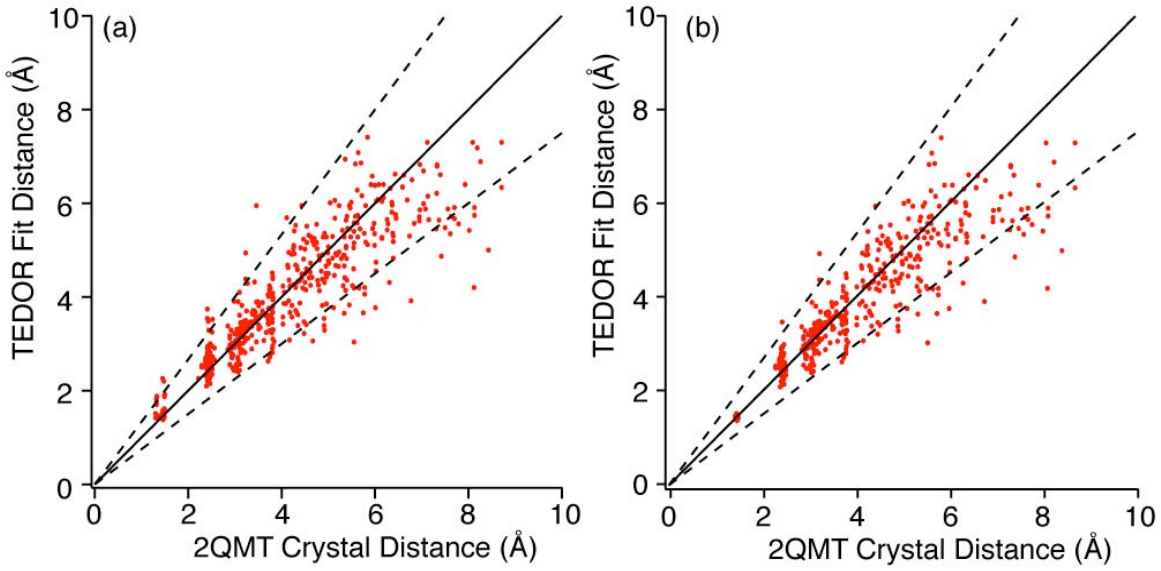

Figure 2.7: Scatter plot of ^{15}N-^{13}C distances determined by TEDOR experiments and X-ray crystallography .

(a) Plot including all ^{15}N sites. (b) Plot including only backbone amide sites. The solid line has a slope of one, to guide the eye, and the dotted lines indicate variations of ±25% from the nominal distance.

We identified and removed intermolecular correlations based on similarity to contacts observed previously.[5, 30, 49] The remaining TEDOR-determined ^{13}C-^{15}N distances showed good agreement with the X-ray structure (Figure 2.7). Excluding the prochiral methyl carbons with ambiguous assignments and sidechain ^{15}N sites, 240 out of 327 distances (73%) determined in the [2]-GB1 sample were within 10% of the X-ray determined distances, and 313 out of 327 (96%) were within 25% of the X-ray distance. For [1,3]-GB1, likewise, 168 out of 260 (65%) were within 10%, and 235 (90%) were within 25%. There is a trend evident in the plot, that outlying points with shorter-than-expected NMR distances are observed more frequently than longer-than-expected NMR distances; we attribute this to the fact that the shorter-than-expected NMR distances are derived from stronger-than-expected cross peaks, which are above the noise floor, whereas the weaker-than-expected cross peaks are not observed at all. In addition,

variations between the TEDOR and X-ray distances are expected because the sample preparations are not identical. For example, we previously found significant differences in sidechain conformation due to subtle differences in crystal polymorphism[45] which more detailed NMR studies could elucidate.

2.6 Protein Structure Calculations with Heteronuclear Distance Restraints

Table 2.1: TEDOR Restraints Used in Structure Calculations

Restraint Type	Never Violate[d]	Violate Once[e]	Violate Twice[f]	Total
Long[a]	108	23	4	135
Medium[b]	53	4	0	57
Sequential[c]	211	17	0	228
Intraresidue	268	42	1	311
Total	640	86	5	731
Percent of Total	87.6	11.8	0.7	

[a] $|i-j| > 4$

[b] $|i-j| \leq 3$

[c] $|i-j| = 1$

[d] Restraints that did not violate by more than 0.3 Å in a majority of the lowest energy models

[e] Restraints that violated once and had 1 Å added to their error.

[f] Restraints that violated after being lengthened and were removed from additional calculations.

We next calculated a *de novo* structure of GB1 in XPLOR-NIH[50] using as experimental restraints the complete set of TEDOR-determined ^{15}N-^{13}C distances. Experimental uncertainties were assumed to have a minimum of 0.1 Å, even in cases where MINUIT reported smaller uncertainties. In addition to distances, TALOS[21] was used to determined backbone dihedral angle from the isotropic chemical shifts. Initial structures calculated this way converged to a ~0.6 Å backbone RMSD fold of GB1, with approximately 36 ^{15}N-^{13}C distances violating by more than 0.5 Å. In cases where distance restraints consistently violated the majority of lowest energy structures, the uncertainties were increased by 1.0 Å and the calculation repeated. Two restraints, T53N-I6CG1 from [1,3]-GB1 and K10NZ-CG from [2]-GB1 (out of the 36) that continued

to violate, despite this increase in uncertainty, were removed altogether. The backbone RMSD converged to 0.35 Å after this round of calculations. The violation threshold was then reduced to 0.3 Å, and another 52 violating distance restraints were relaxed and three more, W43N-V54CG1 from both [1,3]-GB1 and [2]-GB1 and T54N-I6CG1 from [2]-GB1, were removed (86 in total). The process was repeated until no violations greater than 0.3 Å remained. Table I shows all restraints used in the calculation sorted by the distance in the primary structure. Long-range ($|i-j|>3$), medium range ($1<|i-j|<4$), sequential ($|i-j|=1$) and intraresidue correlations are listed along with the number of violations seen in each category. The majority of all restraints (~88%) never violated in any of the calculations.

The violating distance restraints were predominantly in one of two categories: (1) long-range restraints to ^{13}C or ^{15}N sites on flexible side chains and (2) distances involving ^{13}C nuclei with no ^{15}N within ~4 Å. Examples of the first case are I6CD1, I6CG1/2, L7CG, L7CD1/2, K13CE, K13CD and several Lys NZ resonances. As shown by Ishii and Terao, molecular motion alters the effective dipolar couplings in SSNMR,[51] in a manner that for strong couplings will generally result in larger values for ***r*** determined by $\langle 1/r^3 \rangle$ (the NMR observable) than from $\langle r \rangle$ (the X-ray diffraction observable). In the case of ^{13}C nuclei with no nearby ^{15}N, the y scaling factor is highly uncertain, and therefore the covariance between intensity and coupling in this initial rate regime gives the distance estimations a high uncertainty. These effectively serve as long-range distance restraints of quality similar to NOEs.

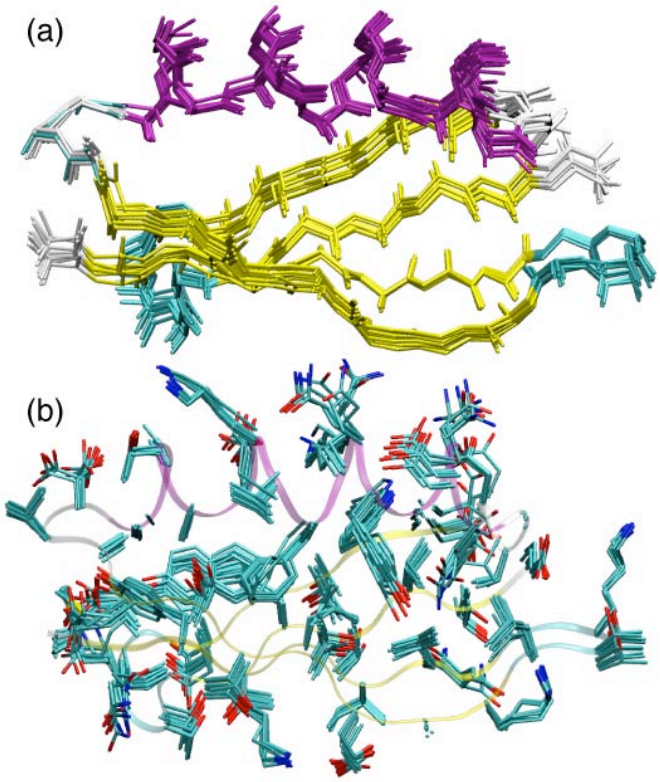

Figure 2.8: GB1 structure calculated using TEDOR distance and TALOS dihedral constraints.

(a) Backbone trace of the 10 lowest energy structures (PDB ID: 2KQ4) out of 260 calculated. (b) Sidechains drawn in CPK coloring scheme. The backbone RMSD is 0.25±0.09 Å within the family of 10 best NMR structures, and the RMSD in comparison to the crystal structure (2QMT) is 0.76±0.06 Å. The all heavy atom RMSD is 0.79±0.03 Å.

Table 2.2: XPLOR Calculation Energies for 10 Lowest Energy Structures

Energies	Average ($kcalmol^{-1}$)	Stdev
Total	228.78	3.52
Bonds	19.35	1.18
Angles	80.15	2.28
VDW	0.03	0.09
TEDOR	115.96	2.93
Dihedral	4.56	0.55
Improper	8.72	0.76

Once violating distance restraints were removed by this procedure, the resulting structure calculation converged to a family of structures (PDB ID: 2KQ4), the 10 lowest

energy out of 260 calculated, with a backbone RMSD of 0.25 ± 0.09 Å with an all atom RMSD of 0.76 ± 0.06 Å. The average structure showed good agreement with the most closely related crystal structure (PDB ID: 2QMT), with a bbRMSD of 0.79 ± 0.03 Å. Figure 2.8 shows the family of best structures shown with (a) only backbone atoms and (b) all side chain atoms. Most variation among the structures is observed in the turns at the beginning and end of the helix. The side chains of residues in the core of the protein are very well constrained leading to very high similarity as indicated by the low all atom RMSD. Thus a majority of the variance in the all atom RMSD is from external side chains that are less well constrained. Table II shows the average XPLOR energies for this family of structures, which indicate that a majority of the pseudoenergy in these structures arises from the TEDOR distance constraints and the XPLOR angle restraints.

2.7 Conclusions

We have demonstrated that the TEDOR pulse sequence can be applied to a preparation of GB1 that was uniformly ^{15}N and sparsely ^{13}C labeled. This sparse ^{13}C labeling allows for long-range TEDOR distances to be observed for every ^{13}C site in GB1. Multiple ^{15}N-^{13}C mixing times were used to create dipolar coupling trajectories for carbon and nitrogen sites up to 6 Å apart, and to detect correlations arising from pairs up to 8 Å apart. These trajectories were fit to exact numerical simulations of TEDOR allowing ^{15}N-^{13}C distances to be determined with both precision and accuracy. When incorporated into simulated annealing calculations these distance restraints, combined with dihedral restraints, produced a very high resolution fold of GB1. Substantially fewer distances were required to determine this structure than in solution NMR studies with similar precision, due to the fact that TEDOR short-to-medium distances can be

determined with greater precision and accuracy than ^{1}H-^{1}H NOEs. Because SSNMR has no intrinsic upper molecular weight limit, this approach is promising for the study of larger nanocrystalline proteins, membrane proteins and insoluble aggregates, in cases where high-resolution 2D ^{15}N-^{13}C spectra can be obtained.

2.8 References

1. Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H., "Structure of a protein determined by solid-state magic-angle- spinning NMR spectroscopy." *Nature* **2002,** *420* (6911), 98-102.
2. Lange, A.; Becker, S.; Seidel, K.; Giller, K.; Pongs, O.; Baldus, M., "A concept for rapid protein-structure determination by solid-state NMR spectroscopy." *Angew. Chem. Int. Ed.* **2005,** *44* (14), 2089-2092.
3. Zech, S. G.; Wand, A. J.; McDermott, A. E., "Protein structure determination by high-resolution solid-state NMR spectroscopy: Application to microcrystalline ubiquitin." *J. Am. Chem. Soc.* **2005,** *127*, 8618-8626.
4. Zhou, D. H.; Shea, J. J.; Nieuwkoop, A. J.; Franks, W. T.; Wylie, B. J.; Mullen, C.; Sandoz, D.; Rienstra, C. M., "Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy." *Angewandte Chemie-International Edition* **2007,** *46* (44), 8380-8383.
5. Franks, W. T.; Wylie, B. J.; Schmidt, H. L. F.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M., "Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (12), 4621-4626.
6. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H., "Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core." *Science* **2008,** *319* (5869), 1523-1526.
7. Opella, S. J.; Marassi, F. M., "Structure determination of membrane proteins by NMR spectroscopy." *Chem. Rev.* **2004,** *104* (8), 3587-3606.
8. Gammeren, A. J.; Hulsbergen, F. B.; Hollander, J. G.; de Groot, H. J., "Residual backbone and side-chain 13C and 15N resonance assignments of the intrinsic transmembrane light-harvesting 2 protein complex by solid-state Magic Angle Spinning NMR spectroscopy." *J. Biomol. NMR* **2005,** *31* (4), 279-93.
9. Kobayashi, M.; Matsuki, Y.; Yumen, I.; Fujiwara, T.; Akutsu, H., "Signal assignment and secondary structure analysis of a uniformly [^{13}C,^{15}N]-labeled membrane protein, H+-ATP synthase subunit c, by magic-angle spinning solid-state NMR." *J. Biomol. NMR* **2006,** *36* (4), 279-293.
10. Frericks, H. L.; Zhou, D. H.; Yap, L. L.; Gennis, R. B.; Rienstra, C. M., "Magic-angle spinning solid-state NMR of a 144 kDa membrane protein complex: E-coli cytochrome bo_3 oxidase." *Journal of Biomolecular NMR* **2006,** *36* (1), 55-71.
11. Etzkorn, M.; Martell, S.; Andronesi, O. C.; Seidel, K.; Engelhard, M.; Baldus, M., "Secondary Structure, Dynamics, and Topology of a Seven-Helix Receptor in Native

Membranes, Studied by Solid-State NMR Spectroscopy." *Angew. Chem. Int. Ed.* **2007,** *46* (3), 459-462.
12. Kijac, A. Z.; Li, Y.; Sligar, S. G.; Rienstra, C. M., "Magic-angle spinning solid-state NMR spectroscopy of nanodisc-embedded human CYP3A4." *Biochemistry* **2007,** *46* (48), 13696-13703.
13. Hiller, M.; Higman, V. A.; Jehle, S.; van Rossum, B. J.; Kuhlbrandt, W.; Oschkinat, H., "[2,3-^{13}C]-labeling of aromatic residues-getting a head start in the magic-angle-spinning NMR assignment of membrane proteins." *J. Am. Chem. Soc.* **2008,** *130* (2), 408-409.
14. Li, Y.; Berthold, D. A.; Gennis, R. B.; Rienstra, C. M., "Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy." *Protein Sci.* **2008,** *17* (2), 199-204.
15. Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M., "Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR." *Proc. Natl. Acad. Sci. USA* **2005,** *102* (44), 15871-15876.
16. Tycko, R., "Molecular structure of amyloid fibrils: insights from solid-state NMR." *Q. Rev. Biophys.* **2006,** *39* (1), 1-55.
17. Kloepper, K. D.; Woods, W. S.; Winter, K. A.; George, J. M.; Rienstra, C. M., "Preparation of alpha-synuclein fibrils for solid-state NMR: Expression, purification, and incubation of wild-type and mutant forms." *Protein Expres. Purif.* **2006,** *48* (1), 112-117.
18. Siemer, A. B.; Arnold, A. A.; Ritter, C.; Westfeld, T.; Ernst, M.; Riek, R.; Meier, B. H., "Observation of highly flexible residues in amyloid fibrils of the HET-s prion." *J. Am. Chem. Soc.* **2006,** *128* (40), 13224-13228.
19. Kloepper, K. D.; Hartman, K. L.; Ladror, D. T.; Rienstra, C. M., "Solid-state NMR spectroscopy reveals that water is nonessential to the core structure of alpha-synuclein fibrils." *J. Phys. Chem. B* **2007,** *111* (47), 13353-13356.
20. Varga, K.; Tian, L.; McDermott, A. E., "Solid-state NMR study and assignments of the KcsA potassium ion channel of S. lividans." *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2007,** *1774* (12), 1604-1613.
21. Cornilescu, G.; Delaglio, F.; Bax, A., "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." *J. Biomol. NMR* **1999,** *13* (3), 289-302.
22. Wylie, B. J.; Schwieters, C. D.; Oldfield, E.; Rienstra, C. M., "Protein Structure Refinement Using 13Cα Chemical Shift Tensors." *J. Am. Chem. Soc.* **2009,** *131* (3), 985-992.
23. Gullion, T.; Schaefer, J., "Rotational-echo double-resonance NMR." *J. Magn. Reson.* **1989,** *81*, 196-200.
24. Hing, A.; Vega, S.; Schaefer, J., "Transferred-echo double-resonance NMR." *J. Magn. Reson.* **1992,** *96*, 205-209.
25. Michal, C. A.; Jelinski, L. W., "REDOR 3D: Heteronuclear distance measurements in uniformly labeled and natural abundance solids." *J. Am. Chem. Soc.* **1997,** *119* (38), 9059-9060.
26. Jaroniec, C. P.; Lansing, J. C.; Tounge, B. A.; Belenky, M.; Herzfeld, J.; Griffin, R. G., "Measurement of dipolar couplings in a uniformly C-13,N-15-labeled membrane

protein: Distances between the Schiff base and aspartic acids in the active site of bacteriorhodopsin." *J. Amer. Chem. Soc.* **2001,** *123* (51), 12929-12930.
27. Jaroniec, C. P.; Filip, C.; Griffin, R. G., "3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon-nitrogen distances in uniformly C-13, N-15-labeled solids." *J. Am. Chem. Soc.* **2002,** *124* (36), 10728-10742.
28. Jaroniec, C. P.; MacPhee, C. E.; Astrof, N. S.; Dobson, C. M.; Griffin, R. G., "Molecular Conformation of a Peptide Fragment of Transthyretin in an Amyloid Fibril." *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (26), 16748-16753.
29. Jaroniec, C. P.; MacPhee, C. E.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G., "High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy." *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (3), 711-716.
30. Helmus, J. J.; Nadaud, P. S.; Hofer, N.; Jaroniec, C. P., "Determination of methyl ^{13}C-^{15}N dipolar couplings in peptides and proteins by three-dimensional and four-dimensional magic-angle spinning solid-state NMR spectroscopy." *J. Chem. Phys.* **2008,** *128* (5), 052314.
31. LeMaster, D. M.; Kushlan, D. M., "Dynamical mapping of E-coli thioredoxin via C-13 NMR relaxation analysis." *J. Am. Chem. Soc.* **1996,** *118* (39), 9255-9264.
32. Hong, M., "Determination of Multiple phi-Torsion Angles in Proteins by Selective and Extensive 13C Labeling and Two-Dimensional Solid-State NMR." *J. Magn. Reson.* **1999,** *139*, 389-401.
33. Wylie, B. J.; Sperling, L. J.; Frericks, H. L.; Shah, G. J.; Franks, W. T.; Rienstra, C. M., "Chemical-shift anisotropy measurements of amide and carbonyl resonances in a microcrystalline protein with slow magic-angle spinning NMR spectroscopy." *J. Am. Chem. Soc.* **2007,** *129* (17), 5318-5319.
34. Ishii, Y.; Ashida, J.; Terao, T., "13C-1H dipolar recoupling dynamics in 13C multiple-pulse solid-state NMR." *Chem. Phys. Lett.* **1995,** *246*, 439-445.
35. Bennett, A. E.; Rienstra, C. M.; Griffiths, J. M.; Zhen, W. G.; Lansbury, P. T.; Griffin, R. G., "Homonuclear radio frequency-driven recoupling in rotating solids." *J. Chem. Phys.* **1998,** *108* (22), 9463-9479.
36. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., "Nmrpipe: a Multidimensional Spectral Processing System Based On Unix Pipes." *J. Biomol. NMR* **1995,** *6* (3), 277-293.
37. Veshtort, M.; Griffin, R. G., "SPINEVOLUTION: A powerful tool for the simulation of solid and liquid state NMR experiments." *J. Magn. Reson.* **2006,** *178*, 248-282.
38. Wylie, B. J.; Sperling, L. J.; Rienstra, C. M., "Isotropic chemical shifts in magic-angle spinning NMR spectra of proteins." *Phys. Chem. Chem. Phys.* **2008,** *10* (3), 405-413.
39. Raleigh, D. P.; Levitt, M. H.; Griffin, R. G., "Rotational resonance in solid state NMR." *Chem. Phys. Lett.* **1988,** *146*, 71-76.
40. Levitt, M. H.; Raleigh, D. P.; Creuzet, F.; Griffin, R. G., "Theory and simulations of homonuclear spin pair systems in rotating solids." *J. Chem. Phys.* **1990,** *92* (11), 6347-64.
41. Helmle, M.; Lee, Y. K.; Verdegem, P. J. E.; Feng, X.; Karlsson, T.; Lugtenburg, J.; de Groot, H. J. M.; Levitt, M. H., "Anomalous rotational resonance spectra in magic-angle spinning NMR." *J. Magn. Reson.* **1999,** *140* (2), 379-403.

42. Duma, L.; Hediger, S.; Lesage, A.; Sakellariou, D.; Emsley, L., "Carbon-13 lineshapes in solid-state NMR of labeled compounds. Effects of coherent CSA-dipolar cross-correlation." *J. Magn. Reson.* **2003,** *162* (1), 90-101.
43. Igumenova, T. I.; McDermott, A. E., "Improvement of resolution in solid state NMR spectra with J-decoupling: an analysis of lineshape contributions in uniformly C-13-enriched amino acids and proteins." *J. Magn. Reson.* **2003,** *164* (2), 270-285.
44. Franks, W. T.; Zhou, D. H.; Wylie, B. J.; Money, B. G.; Graesser, D. T.; Frericks, H. L.; Sahota, G.; Rienstra, C. M., "Magic-angle spinning solid-state NMR spectroscopy of the beta-1 immunoglobulin binding domain of protein G (GB1): ^{15}N and ^{13}C chemical shift assignments and conformational analysis." *J. Am. Chem. Soc.* **2005,** *127* (35), 12291-12305.
45. Frericks Schmidt, H. L.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. M., "Crystal polymorphism of protein GB1 examined by solid-state NMR spectroscopy and x-ray diffraction." *J. Phys. Chem. B* **2007,** *111*, 14362-14369.
46. Mueller, K. T.; Jarvie, T. P.; Aurentz, D. J.; Roberts, B. W., "The Redor Transform - Direct Calculation of Internuclear Couplings from Dipolar-Dephasing Nmr Data." *Chem. Phys. Lett.* **1995,** *242* (6), 535-542.
47. Mueller, K. T., "Analytic solutions for the time evolution of dipolar-dephasing NMR signals." *J. Magn. Reson. A* **1995,** *113* (1), 81-93.
48. Schaefer, J., "REDOR-determined distances from heterospins to clusters of C-13 labels." *J. Magn. Reson.* **1999,** *137* (1), 272-275.
49. Peng, X. H.; Libich, D.; Janik, R.; Harauz, G.; Ladizhansky, V., "Dipolar chemical shift correlation spectroscopy for homonuclear carbon distance measurements in proteins in the solid state: Application to structure determination and refinement." *J. Am. Chem. Soc.* **2008,** *130* (1), 359-369.
50. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., "The Xplor-NIH NMR molecular structure determination package." *J. Magn. Reson.* **2003,** *160* (1), 65-73.
51. Ishii, Y.; Terao, T.; Hayashi, S., "Theory and simulation of vibrational effects on structural measurements by solid-state nuclear magnetic resonance." *J. Chem. Phys.* **1997,** *107* (8), 2760-2774.

Chapter 3

Supramolecular Protein Structure Determination by Site-Specific Long-Range Intermolecular Solid State NMR Spectroscopy

3.1 Notes and Acknowledgements

Reproduced in part with permission from Nieuwkoop, A. J. and Rienstra C. M. "Supramolecular Protein Structure Determination by Site-Specific Long-Range Intermolecular Solid State NMR Spectroscopy" Journal of the American Chemistry Society, Vol 131 Issue Copyright 2010 American Chemical Society.

We thank the National Institutes of Health (R01-GM073770, an American Recovery and Reinvestment Act supplement, and the Molecular Biophysics Training Program at the University of Illinois) for financial support, Dr. Charles Schwieters for assistance in the application of XPLOR-NIH in multimeric systems, Lindsay Sperling for assistance with recreating crystal lattice coordinates, Damien Mathew for advice in performing VMD scripting, and the School of Chemical Sciences NMR Facility at the University of Illinois.

3.2 Introduction

Solid-state NMR (SSNMR) spectroscopy is a uniquely effective method for protein structure determination of insoluble protein fibrils or aggregates.[1-4] SSNMR spectra of such samples often contain intramolecular and intermolecular correlations with similar intensities, leading to ambiguities in interpretation of data sets obtained from undiluted, uniformly-^{13}C,^{15}N-labeled samples. In order to faithfully reproduce tertiary and quaternary structural features, differential isotopic labeling strategies must be utilized. For example, intensities of intermolecular crosspeaks are attenuated by diluting the labeled sample in natural abundance material. Alternatively, these signals can be accentuated by utilizing regio-specific ^{15}N and ^{13}C labeling, as demonstrated in studies of reassembled thioredoxin,[5] or recrystallizing physical mixtures of ^{15}N and ^{13}C labeled proteins, such as Crh.[6] With such labeling patterns, heteronuclear distance techniques—such as REDOR,[7] TEDOR,[8] and NHHC methods[6]—can be applied to obtain site-specific distance restraints across molecular interfaces.

Here we extend this approach to the full, atomic-resolution structure determination of a quaternary protein assembly in the nanocrystalline state. We utilize 3D Z-filtered TEDOR,[9] as recently demonstrated to quantify intramolecular distances in proteins,[10] to detect site-resolved intermolecular correlations. The distance restraints are incorporated into simulated annealing calculations, resulting in a specific quaternary arrangement that is independent of the initial condition. For example, if the coordinates from the known trigonal X-ray crystal structure (PDB ID 2QMT)[11] are used as the starting point for the calculation, the NMR restraints refine the exact atom positions but do not change the quaternary arrangement; if instead the orthorhombic X-ray lattice coordinates are

used (PDB ID 2GI9),[12] in the course of the calculation, the relative orientations of neighboring molecules change substantially, transforming into the trigonal form. Finally, calculations initialized with isolated monomers converge to the trigonal form. These results demonstrate that SSNMR methods can fully reproduce not only secondary and tertiary, but also quaternary, structural features at atomic-resolution detail.

3.3 Methods

3.3.1 Sample Preparation

Samples of GB1 were prepared by bacterial overexpression in two batches: (a) ^{15}N isotopically labeled in media containing ^{15}N ammonium chloride and natural abundance glucose, and (b) ^{13}C isotopically labeled in 1,3-^{13}C-glycerol and natural abundance sodium carbonate and ammonium chloride. The sample was prepared by physically mixing a 50:50 ratio of these two preparations prior to nanocrystalline precipitation by previously reported methods. Additional samples of ^{13}C and ^{15}N enriched GB1 were also prepared in media containing ^{15}N ammonium chloride and either 2-^{13}C-glycerol and calcium ^{13}C-carbonate as reported earlier.[1] Samples (each with ~18 mg protein) were packed into the central 80% of limited speed 3.2 mm Varian rotors (Varian, Inc., Fort Collins, Colorado).

3.3.2 Solid-state NMR spectroscopy

Experiments were performed using a 500 MHz InfinityPlus spectrometer (Varian, Inc., Palo Alto, California and Fort Collins, Colorado) equipped with a 3.2 mm T3 Balun^{TM} ^{1}H-^{13}C-^{15}N MAS probe. Pulse widths (π/2) for ^{1}H, ^{13}C, and ^{15}N were 2.55 μs, 2.9 μs, and 5.5 μs respectively. Spinning was controlled with a Varian MAS controller to 11,111 ± 2 Hz. TPPM decopuling[13] (~75 kHz, ^{1}H pulse width 6.6 μs, total phase

difference 15°) was used during acquisition and ^{15}N evolution. TPPM during REDOR periods was independently optimized at a higher power level (~100 kHz, 5 μs, 11°).

ZF-TEDOR 2D planes with 8.64, 16.86 and 21.6 ms of mixing were acquired. Each was digitized to 3072 x 10 μs in the direct (^{13}C) and 80 x 180 μs in the indirect (^{15}N) dimension. Total measurement times were 21.7, 34.6 and 26 hrs.

3.3.3 Structure Calculation Methods

Simulated annealing calculations were performed using the python extension of XPLOR-NIH.[14] Intermolecular TEDOR restraints were applied ambiguously from the central GB1 molecule to any other GB1, enforcing the intermolecular nature of the data. Intramolecular restraints were applied to each monomer and included TEDOR distances and TALOS dihedral angle restraints. A symmetry potential was used to enforce similarity between all monomers in each calculation, as required by our NMR data which shows only a single set of peaks for each residue. For the docking procedure this potential was also used to fix the internal coordinates of the monomers: by fixing the atom positions of the central monomer and applying a large symmetry potential the center of mass and orientation of the peripheral GB1 monomers were allowed to change while keeping the internal structure fixed. High temperature dynamic times of 20 ps were used for the trigonal and in silico calculations and 40 ps for the orthorhombic to trigonal annealing followed by annealing times of 60 ps. In all cases 600 structures were calculated with the set of lowest 10 energy structures reported. bbRMSDs reported for these sets are the average of the 10 RMSDs for the backbone aligned in VMD-XPLOR, reported error is the standard deviation. In the cases where bbRMSDs are reported to

crystal structures the same structures are aligned to the X-ray crystal lattice in VMD-XPLOR[15] and the RMSD of all 10 structures, to the X-ray was reported.

3.4 ZF-TEDOR on Mixed Labeled Samples

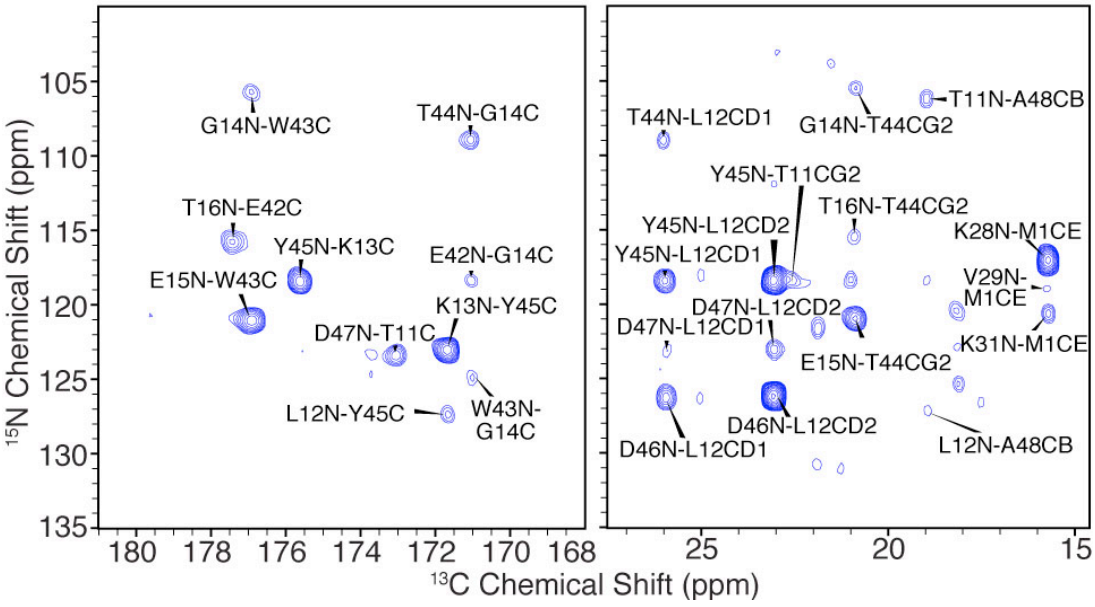

Figure 3.1. 2D ^{15}N-^{13}C plane of 3D ZF-TEDOR spectrum.

ZF-TEDOR plane illustrating intermolecular correlations from a recrystallized physical mixture (50:50) of 1,3-^{13}C-glycerol and ^{15}N-labeled GB1. Data were acquired at 500 MHz (^{1}H) with 21.6 ms ^{15}N-^{13}C mixing, 14.4 ms ^{15}N and 30.7 ms ^{13}C evolution. Total measurement time 26 hrs.

Previously we have collected a range of structural data for GB1 in the nanocrystalline state—including homonuclear distance restraints and vector angles,[16] backbone chemical shift tensors,[17] and high-precision ^{15}N-^{13}C distance restraints.[10] The 3D Z-filtered TEDOR pulse sequence produces hundreds of intramolecular ^{15}N-^{13}C restraints, which are sufficient (with TALOS[18] restraints) to define an atomic resolution protein structure (PDB ID 2KQ4).[10] Extending this approach to physical mixtures (50:50) of ^{13}C- and ^{15}N-labeled molecules results in a rich set of intermolecular restraints. The spectra were collected with a 500 MHz Varian InfinityPlus spectrometer and 3.2-mm

BalunTM ^{1}H-^{13}C-^{15}N probe, with ~100 kHz TPPM decoupling during the REDOR periods. Under these conditions, the ^{15}N and ^{13}C (especially methyl and carbonyl) T2 values are >25 ms, coherence lifetimes that in combination with the sparse labeling pattern (mitigating effects of scalar ^{13}C-^{13}C couplings) enabled the detection of particularly strong TEDOR cross peaks. The peak intensities continue to increase up to mixing times of at least 20 ms (Figure 3.1). Most of these correlations are intermolecular correlations (natural abundance, intramolecular correlations are observed but are much weaker, and can be readily identified with spectra at short mixing times, (Figure 3.3 below). Reverse labeling of the ^{15}N sample with ^{13}C-depleted glucose would further suppress these undesirable peaks, and this strategy would be beneficial for proteins of higher molecular weight and/or greater spectral degeneracy.

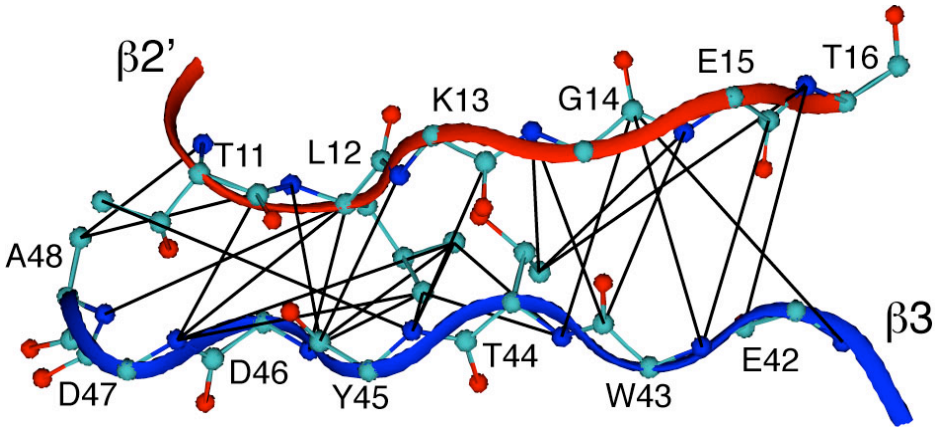

Figure 3.2: Intermolecular ^{15}N-^{13}C distance restraints observed by TEDOR.

The β2'-β3 intermolecular interface of GB1 nanocrystals with intermolecular restraints, with maximum distances of 5, 7 and 8 Å, determined by the TEDOR mixing time at which the peaks were first observed.

The observed ^{15}N and ^{13}C linewidths were ~0.5 and 0.2-0.3 ppm, respectively, enabling unique (unambiguous) assignment of several intermolecular correlations

based on the published GB1 chemical shifts.[19] From these assignments, we identified two distinct molecular interfaces: (1) an anti-parallel intermolecular beta-sheet (b2'-b3), defined by long-range correlations with (i+j) = 58 ± 2 (Figure 3.2) and (2) contacts between helical residues 28-31 and residues 1 and 20-21. A subset of these correlations has been previously reported.[10, 16, 20, 21] We obtained a sufficiently large number of unambiguous restraints that structure calculations generally converged well and clarified the assignment of ambiguous restraints.

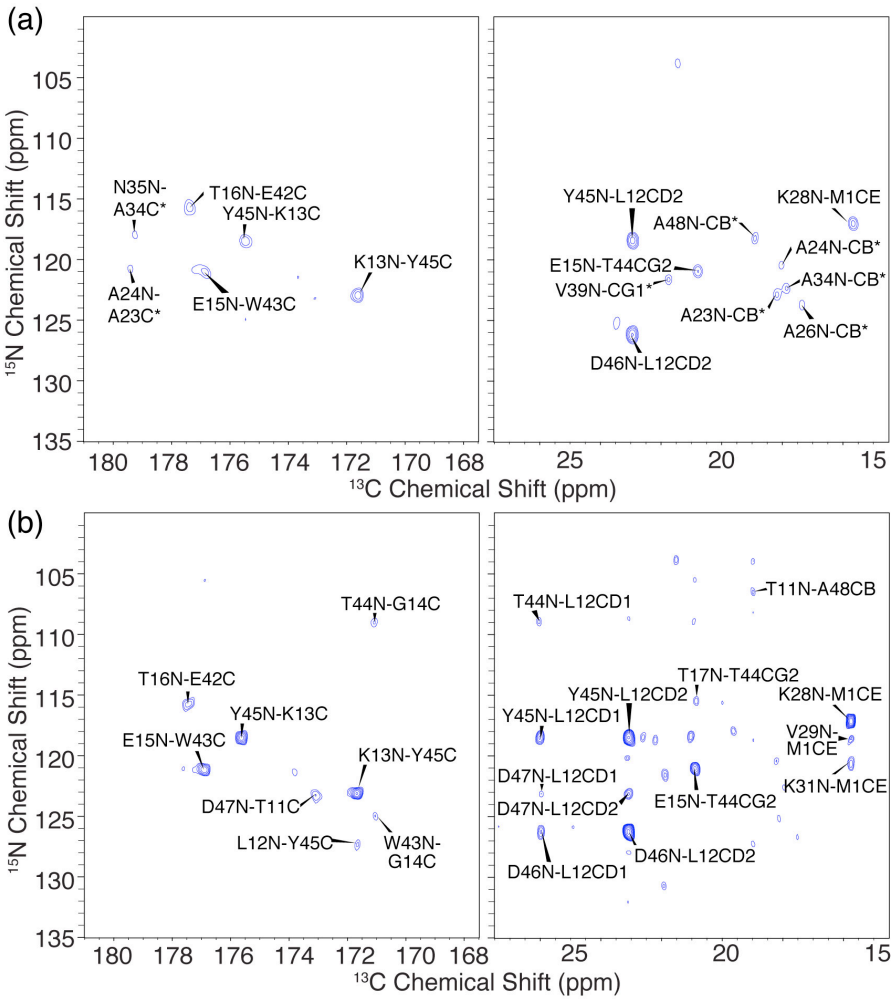

Figure 3.3. 2D ^{13}C-^{15}N planes from the 3D ZF-TEDOR spectrum collected with 50% U-^{15}N GB1 physically mixed with 50% 1,3-^{13}C glycerol labeled GB1.

TEDOR mixing times of (a) 8.64 ms and (b) 16.48 ms are shown. Spectra were acquired at 500 MHz (^{1}H) with 21.6 ms ^{15}N-^{13}C mixing, 14.4 ms t_{1max} (^{15}N evolution) and 30.7 ms ^{13}C acquisition time. Total measurement time was (a) 21.7 and (b) 34.6 hours. Peaks in (a) labeled with asterisks are attributed to natural abundance ^{13}C in the ^{15}N labeled samples. These weak peaks decrease in intensity at longer than 8 ms mixing times and therefore are concluded to be intramolecular and were not included in the restraint lists for the structure calculations.

Intermolecular TEDOR restraints were extracted from 2D TEDOR planes by first picking peaks from the longest (21.6 ms, Figure 3.1) ^{15}N-^{13}C mixing experiments and

then determining which of these peaks also appeared in the shorter mixing experiments (Figure 3.3). A small subset of peaks—which we attribute to natural abundance ^{13}C sites on the ^{15}N labeled sample—appear at the shortest mixing time; however, these did not prove difficult to differentiate from the labeled sites. Specifically, the maximum ^{13}C-^{15}N transfer for TEDOR is theoretically 52%, and we observe on the order of 40% for 1-bond (1.45 Å, at 1.44 ms) and 30% for a 2-bond (2.45 Å, at 5.76 ms) ^{15}N-^{13}C distances. Thus the strongest natural abundance peaks will appear at 0.4% and 0.3% respectively of the reference spectrum, and these intensities decay at longer mixing times. For example, the natural abundance peaks observed (Figure 3.3a, labeled with asterisks) decay below the noise threshold at longer mixing times (Figure 3.3b and Figure 3.1). At mixing times of ~20 ms, the natural abundance ^{13}C sites decay to a small fraction of their initial intensity, and the intermolecular correlations increase in intensity.

Intermolecular TEDOR restraints were assumed to have distance ranges <5 Å for restraints first appearing at 8.64 ms mixing, <7 Å for peaks that were first observed in the 16.86 ms data and <8 Å for those peaks only seen in the 21.6 ms mixing time spectrum. These distances correspond to the longest-range correlations observed in the undiluted 15N-13C sample[10] at moderate (~8 ms, 6 Å) and long (~15 ms, 8 Å) mixing times, corrected for the relative amount of sample and signal averaging time; i.e., with 25% of the sample quantity (50% is labeled with either ^{15}N or ^{13}C, and only 50% of intermolecular interfaces have the complementary labeling to enable ^{15}N-^{13}C correlations) but twice the signal averaging time, we observed a ~2-fold decrease in signal to noise. Therefore we adjusted the distance threshold (assuming the linear rate regime of TEDOR buildup) by 2^(-1/3) = ~0.8.

3.5 In silico Annealing of the GB1 Crystal Lattice

Correlations were assigned in a manner explicitly taking ambiguity into account; possible assignments included any ^{15}N resonance within 0.25 ppm and any ^{13}C resonance within 0.15 ppm of the observed crosspeak frequency (corresponding to half the average linewidths in each dimension). This procedure resulted in restraints with up to four-fold ambiguity in each dimension. For example, the peak at 123.3 ppm 15N, and 173.1 ppm 13C (Figures 3.1 and 3.3) has nine possible assignments (three-fold degenerate in each dimension): Y3N, K13N or D47N correlated to K4C, G9C or T11C. However, 20 restraints were completely unambiguous, while 18 were two-fold and 14 were threefold degenerate. The remaining 36 were four or more fold degenerate (all intermolecular restraints are shown in Table 3.1 at the end of the Chapter).

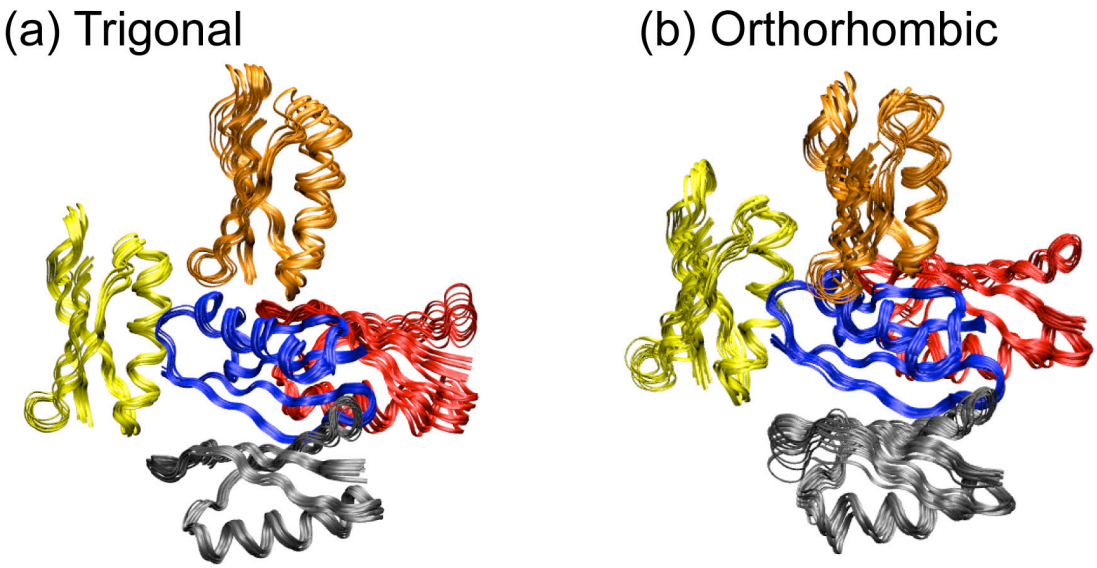

Figure 3.4. Quaternary structure of GB1 (10 lowest energy structures out of 600) as determined by simulated annealing structure calculations.

The 10 lowest energy structures out of 600 calculated, initiated from the (a) trigonal (PDB ID 2QMT) and (b) orthorhombic (PDB ID 2GI9) crystal lattice. Ambiguous intermolecular TEDOR restraints were used for which all possible 15N and 13C assignments for a given correlation were included.

The crystal lattices used as the starting structures for the trigonal (2QMT) and orthorhombic (2GI9) calculations were generated using SwissPDB viewer[22] (http://www.expasy.org/spdbv/) selecting all GB1 monomers with ^{15}N sites within 8 Å of a ^{13}C site on the central monomer. In both cases a total of six GB1 molecules were found to have ^{15}N-^{13}C pairs less than 8 Å apart. The calculation was performed starting with each crystal form, applying identical restraints. Intermolecular restraints were assigned taking into account the inherent ambiguities; for each peak, all ^{15}N and ^{13}C resonance frequencies within half the peak linewidth were deemed possible assignments. The ambiguous restraints were then formalized in XPLOR-NIH lists. Figure 3.4 illustrates the results for calculations initiated using the (a) trigonal and (b) orthorhombic crystal packing geometries. In the case of the trigonal packing one of those monomers was only restrained by one correlation, and as such did not converge to a unique conformation in the final calculations. Thus in Figure 3.4 only the five well-restrained monomers are shown. Similarly for the orthorhombic lattice the sixth monomer was not restrained at all and is not shown in Figure 3.4.

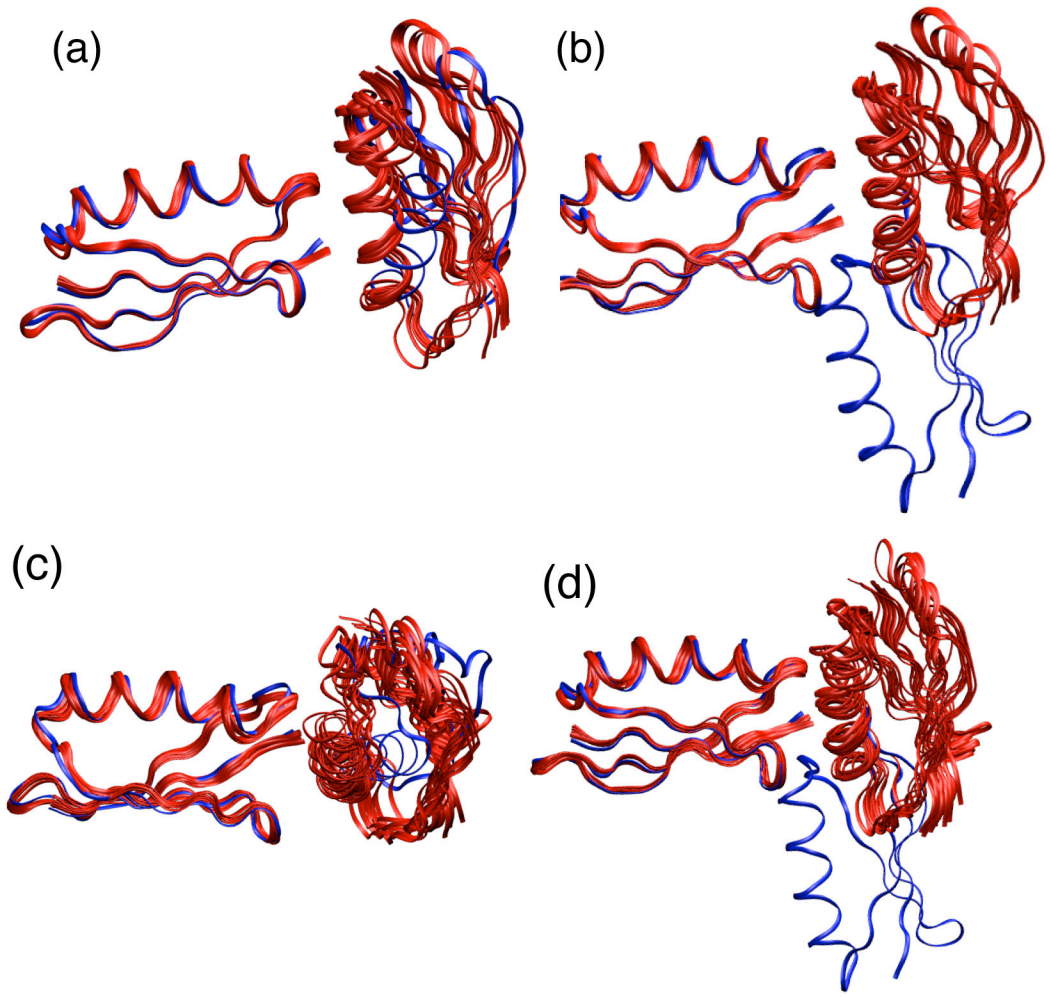

Figure 3.5 Agreement of intermolecular TEDOR structures with initial crystal lattices.

Two monomers (10 lowest energy out of 600 calculated) taken from the (a,b) trigonal and (c,d) orthorhombic calculations with intermolecular TEDOR restraint lists. These two monomers (red) aligned by the central monomer overlaid with (a,c) the trigonal (2QMT, blue) and (b,d) the orthorhombic (2GI9, blue) X-ray crystal lattices.

In both the calculation starting trigonal and orthorhombic the final structure agreed well with the trigonal crystal lattice. Figure 3.5 shows the agreement for two monomers most distinct between the orthorhombic and trigonal forms. Backbone RMSDs to the trigonal lattice are 4.1 ± 0.4 Å for the structure that started from the trigonal lattice (Figure 3.5a) and 4.7 ± 0.4 Å for the structure starting with the orthorhombic lattice

(Figure 3.5c). Conversely, backbone RMSDs to the orthorhombic lattice were 13.6 Å ± 0.3 Å for the structure that started from the trigonal lattice (Figure 3.5b) and 15.8 Å ± 0.7 Å for the structure starting with the orthorhombic lattice (Figure 3.5d). This demonstrates that the intermolecular TEDOR data is sufficient to differentiate between the trigonal and orthorhombic crystal lattices regardless of which lattice the calculation is initiated from.

3.6 De novo Determination of the GB1 Crystal Lattice

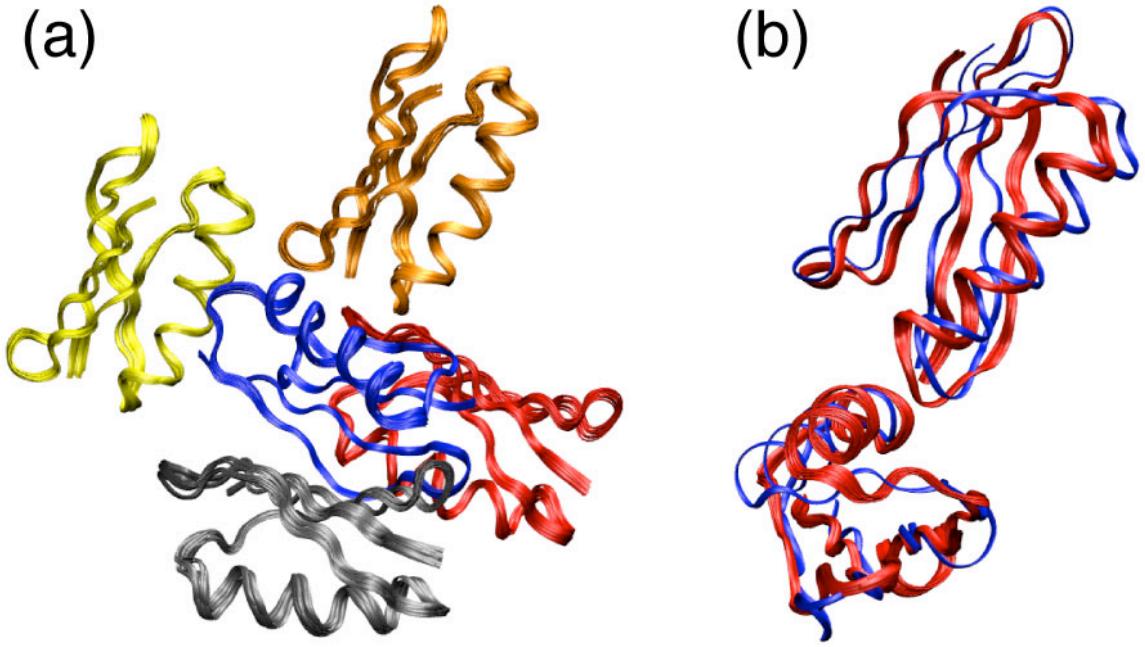

Figure 3.6: The crystal packing of GB1 determined by simulated annealing of five GB1 monomers with ambiguous intermolecular TEDOR restraints.

The bbRMSD of the 10 lowest energy structures (a) is 0.42 ± 0.08 Å. Alignment of the region distinct to the trigonal lattice (b) yields a 2.9 ± 0.1 Å bbRMSD of the ensemble (red) to the 2QMT crystal structure (blue).

Given the success of the intermolecular TEDOR restraints in annealing possible crystal forms to the correct trigonal lattice, we next attempted a *de novo* lattice determination. For this *de novo* annealing calculation five copies of the lowest energy model from the intramolecular TEDOR structure (2KQ4) were manually translated and rotated into a grid with four monomers in plane and one above them out of plane. VMD-XPLOR[15] (http://vmd-xplor.cit.nih.gov/) was used to create the initial conditions for the in silico annealing. All monomers were ~20 Å apart, and positioned in a rough array around the central GB1, corresponding to the four areas of intermolecular contact observed. The calculations proved to be relatively insensitive to starting conditions, converging to a similar structure from multiple starting orientations, with the only requirement being all monomers must be separated by an equal distance from the central monomer. Intermolecular restraints were first applied with internal coordinates of the monomers held constant, allowing the intermolecular TEDOR restraints to dock the peripheral monomers to the central monomer. If docking was successful, defined by a lack of VDW clashes and/or intercalated monomers, the structure was then refined. The resulting quaternary arrangements (Figure 3.6a) showed high precision (bbRMSD for all 5 monomers of <0.5 Å) and accuracy (Figure 3.6b), with a 5 Å bbRMSD to the trigonal crystal structure (2QMT), but a bbRMSD of more than 13 Å from the orthorhombic structure (2GI9). Hence, this is a de novo recreation of the crystal lattice of GB1.

3.7 Conclusions

We have demonstrated that 3D Z-filtered TEDOR experiments, when performed on mixtures of isotopically labeled protein samples, report on site-specific intermolecular distance restraints. These data sets can be leveraged to perform rigorous structure

calculations of the protein interface. In the example demonstrated here, we have determined the packing arrangement of our nanocrystalline GB1 preparation to be consistent with the trigonal form as determined by X-ray diffraction. Therefore this represents an important proof of principle, in a case where the results can be directly compared with other structural information. We envision the application of this approach to determining the registry and quaternary arrangement of protein fibrils, which most often cannot be determined by diffraction methods.

Table 3.1 Intermolecular TEDOR Correlations

Residue/Atom	Residue/Atom	Distance (Å)	Degeneracy	Manual Assignment
1CE	31N	5	1	K31N-M1CE
31N	1CE	5	1	K31N-M1CE
20CB	31NZ	8	1	K31NZ-A20CB
31NZ	20CB	8	1	K31NZ-A20CB
1CB	31NZ	5	1	K31NZ-M1CB
31NZ	1CB	5	1	K31NZ-M1CB
1CE	31NZ	5	1	K31NZ-M1CE
31NZ	1CE	5	1	K31NZ-M1CE
21CG1	31NZ	5	1	K31NZ-V21CG1
31NZ	21CG1	5	1	K31NZ-V21CG1
12N	48CB	5	1	L12N-A48CB
48CB	12N	5	1	L12N-A48CB
1N	27CG	7	1	M1N-A27CG
27CG	1N	7	1	M1N-A27CG
1N	27CD	5	1	M1N-E27CD
27CD	1N	5	1	M1N-E27CD
11N	48CB	8	1	T11N-A48CB
48CB	11N	8	1	T11N-A48CB
12CD2	44N	7	1	T44N-L12CD2
44N	12CD2	7	1	T44N-L12CD2
20N	31CG or 31CG	7	2	A20N-K31CG
31CG or 6CG1	20N	7	2	A20N-K31CG
46N or 6N	12CD2	5	2	D46N-L12CD2
12CD2	46N or 6N	5	2	D46N-L12CD2
14N	44CG2 or 29CG1	5	2	G14N-T44CG2
44CG2 or 29CG1	14N	5	2	G14N-T44CG2
14N	43C or 26C	5	2	G14N-W43C
43C or 26C	14N	5	2	G14N-W43C
1CE	28N or 25N	5	2	K28N-M1CE

Table 3.1, continued

Residue/Atom	Residue/Atom	Distance (Å)	Degeneracy	Manual Assignment
28N or 25N	1CE	5	2	K28N-M1CE
12N	45C or 16C	5	2	L12N-Y45C
45C or 16C	12N	5	2	L12N-Y45C
1N	24CB or 23CB	8	2	M1N-A24CB
24CB or 23CB	1N	8	2	M1N-A24CB
12CD2	53N or 51N	7	2	T53N-L12CD2
53N or 51N	12CD2	7	2	T53N-L12CD2
15C	43N or 8N	8	2	W43N-E15C
43N or 8N	15C	8	2	W43N-E15C
46N	12CD1 or 7CD1 or 13CG	5	3	D46N-L12CD1
12CD2	47N or 3N or 13N	5	3	D47N-L12CD2
47N or 3N or 13N	12CD2	5	3	D47N-L12CD2
31NZ	20C or 42C or 32C	7	3	K31NZ-A20C
20C or 42C or 32C	31NZ	7	3	K31NZ-A20C
44N	14C or 18C or 1C	5	3	T44N-G14C
14C or 18C or 1C	44N	5	3	T44N-G14C
44N	7CD1 or 13CG or 12CD1	5	3	T44N-L12CD1
12CD1 or 7CD1 or 13CG	44N	5	3	T44N-L12CD1
12CD2	45N or 30N or 54N	5	3	Y45N-L12CD2
45N or 30N or 54N	12CD2	5	3	Y45N-L12CD2
11CG2	45N or 30N or 54N	7	3	Y45N-T11CG2
11CG2	45N or 30N or 54N	7	3	Y45N-T11CG2
45N or 30N or 54N	11CG2	7	3	Y45N-T11CG2
1CE	42N or 45N or 48N or 30N	7	4	F30N-M1CE
45N or 30N or 42N or 48N	1CE	7	4	F30N-M1CE
13N or 47N	45C or 16C	5	4	K13N-Y45C
45C or 16C	13N or 47N	5	4	K13N-Y45C
31NZ	21C or 3C or 40C or 2C	5	4	K31NZ-V21C
21C or 3C or 2C or 40C	31NZ	5	4	K31NZ-V21C
19N or 2N	24CB or 23CB	7	4	Q2N-A23CB
24CB or 23CB	2N or 19N	8	4	Q2N-A23CB
46N or 6N	12CD1 or 7CD1 or 13CG	5	6	D46N-L12CD1
12CD1 or 7CD1 or 13CG	46N or 6N	5	6	D46N-L12CD1
36CB or 8CB	32NE2 or 17N or 22N	7	6	Q32NE2-D36CB
32NE2 or 17N or 22N	36CB or 8CB	7	6	Q32NE2-D36CB
44CG2 or 29CG1	16N or 22N or 32NE2	7	6	T16N-T44CG2
16N or 22N or 32NE2	44CG2 or 29CG1	7	6	T16N-T44CG2
43N or 8N	14C or 18C or 1C	5	6	W43N-G14C
14C or 18C or 1C	43N or 8N	5	6	W43N-G14C
46N or 6N	12CA or 47CA or 24CA or 23CA	5	8	D46N-L12CA
12CA or 47CA or 24CA or 23CA	46N or 6N	5	8	D46N-L12CA
44CG2 or 29CG1	15N or 32N or 36N or 10N	5	8	E15N-T44CG2
15N or 32N or 36N or 10N	44CG2 or 29CG1	5	8	E15N-T44CG2

Table 3.1, continued

Residue/Atom	Residue/Atom	Distance (Å)	Degeneracy	Manual Assignment
43C or 26C	15N or 32N or 36N or 10N	5	8	E15N-W43C
15N or 32N or 36N or 10N	43C or 26C	5	8	E15N-W43C
12CD1 or 7CD1 or 13CG	47N or 3N or 13N	7	9	D47N-L12CD1
47N or 3N or 13N	12CD1 or 7CD1 or 13CG	7	9	D47N-L12CD1
11C or 9C or 4C	47N or 3N or 13N	5	9	D47N-T11C
47N or 3N or 13N	9C or 4C or 11C	5	9	D47N-T11C
12CD1 or 7CD1 or 13CG	45N or 30N or 54N	5	9	Y45N-L12CD1
45N or 30N or 54N	7CD1 or 13CGor 12CD1	5	9	Y45N-L12CD1
47N or 3N or 13N	12CA or 47CA or 24CA or 23CA	5	12	D47N-L12CA
12CA or 47CA or 24CA or 23CA	47N or 3N or 13N	5	12	D47N-L12CA
14C or 18C or 1C	42N or 45N or 48N or 30N	8	12	E42N-G14C
45N or 30N or 42N or 48N	14C or 18C or 1C	8	12	E42N-G14C
19C or 25C or 13C or 49C or 52C	45N or 30N or 54N	5	15	Y45N-K13C
45N or 30N or 54N	19C or 25C or 13C or 49C or 52C	5	15	Y45N-K13C
12CA or 47CA or 24CA or 23CA	45N or 30N or 42Nor 48N	5	16	A48N-L12CA
45N or 30N or 42N or 48N	12CA or 47CA or 24CAor 23CA	5	16	A48N-L12CA

3.8 References

1. Petkova, A. T.; Leapman, R. D.; Guo, Z. H.; Yau, W. M.; Mattson, M. P.; Tycko, R., "Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils." *Science* **2005,** *307* (5707), 262-265.
2. Tycko, R., "Molecular structure of amyloid fibrils: insights from solid-state NMR." *Q. Rev. Biophys.* **2006,** *39* (1), 1-55.
3. Helmus, J. J.; Surewicz, K.; Nadaud, P. S.; Surewicz, W. K.; Jaroniec, C. P., "Molecular conformation and dynamics of the Y145Stop variant of human prion protein." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (17), 6284-6289.
4. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H., "Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core." *Science* **2008,** *319* (5869), 1523-1526.
5. Yang, J.; Tasayco, M. L.; Polenova, T., "Magic angle spinning NMR experiments for structural studies of differentially enriched protein interfaces and protein assemblies." *J. Am. Chem. Soc.* **2008,** *130* (17), 5798-5807.

6. Etzkorn, M.; Bockmann, A.; Lange, A.; Baldus, M., "Probing molecular interfaces using 2D magic-angle-spinning NMR on protein mixtures with different uniform labeling." *J. Am. Chem. Soc.* **2004,** *126* (45), 14746-14751.
7. Gullion, T.; Schaefer, J., "Rotational-echo double-resonance NMR." *J. Magn. Reson.* **1989,** *81*, 196-200.
8. Hing, A.; Vega, S.; Schaefer, J., "Transferred-echo double-resonance NMR." *J. Magn. Reson.* **1992,** *96*, 205-209.
9. Jaroniec, C. P.; Filip, C.; Griffin, R. G., "3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon-nitrogen distances in uniformly C-13, N-15-labeled solids." *J. Am. Chem. Soc.* **2002,** *124* (36), 10728-10742.
10. Nieuwkoop, A. J.; Wylie, B. J.; Franks, W. T.; Shah, G. J.; Rienstra, C. M., "Atomic resolution protein structure determination by three-dimensional transferred echo double resonance solid-state nuclear magnetic resonance spectroscopy." *J. Chem. Phys.* **2009,** *131* (9), 095101.
11. Frericks Schmidt, H. L.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. M., "Crystal polymorphism of protein GB1 examined by solid-state NMR spectroscopy and x-ray diffraction." *J. Phys. Chem. B* **2007,** *111*, 14362-14369.
12. Franks, W. T.; Wylie, B. J.; Stellfox, S. A.; Rienstra, C. M., "Backbone conformational constraints in a microcrystalline U-^{15}N-labeled protein by 3D dipolar-shift solid-state NMR spectroscopy." *J. Am. Chem. Soc.* **2006,** *128* (10), 3154-3155.
13. Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G., "Heteronuclear decoupling in rotating solids." *J. Chem. Phys.* **1995,** *103* (16), 6951-6958.
14. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., "The Xplor-NIH NMR molecular structure determination package." *J. Magn. Reson.* **2003,** *160* (1), 65-73.
15. Schwieters, C. D.; Clore, G. M., "The VMD-XPLOR visualization package for NMR structure refinement." *J. Magn. Reson.* **2001,** *149* (2), 239-244.
16. Franks, W. T.; Wylie, B. J.; Schmidt, H. L. F.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M., "Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (12), 4621-4626.
17. Wylie, B. J.; Schwieters, C. D.; Oldfield, E.; Rienstra, C. M., "Protein Structure Refinement Using 13Cα Chemical Shift Tensors." *J. Am. Chem. Soc.* **2009,** *131* (3), 985-992.
18. Cornilescu, G.; Delaglio, F.; Bax, A., "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." *J. Biomol. NMR* **1999,** *13* (3), 289-302.
19. Franks, W. T.; Zhou, D. H.; Wylie, B. J.; Money, B. G.; Graesser, D. T.; Frericks, H. L.; Sahota, G.; Rienstra, C. M., "Magic-angle spinning solid-state NMR spectroscopy of the beta-1 immunoglobulin binding domain of protein G (GB1): ^{15}N and ^{13}C chemical shift assignments and conformational analysis." *J. Am. Chem. Soc.* **2005,** *127* (35), 12291-12305.
20. Helmus, J. J.; Nadaud, P. S.; Hofer, N.; Jaroniec, C. P., "Determination of methyl ^{13}C-^{15}N dipolar couplings in peptides and proteins by three-dimensional and four-

dimensional magic-angle spinning solid-state NMR spectroscopy." *J. Chem. Phys.* **2008,** *128* (5), 052314.
21. Peng, X. H.; Libich, D.; Janik, R.; Harauz, G.; Ladizhansky, V., "Dipolar chemical shift correlation spectroscopy for homonuclear carbon distance measurements in proteins in the solid state: Application to structure determination and refinement." *J. Am. Chem. Soc.* **2008,** *130* (1), 359-369.
22. Guex, N.; Peitsch, M. C., "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." *Electrophoresis* **1997,** *18* (15), 2714-23.

Chapter 4

Proton Detected SSNMR Spectroscopy of Alpha-Synuclein Fibrils

4.1 Notes and Acknowledgments.

This chapter is adapted from "Long-range distance restraints in alpha synuclein fibrils by proton detected SSNMR spectroscopy of $_1$H diluted samples." Nieuwkoop, A. J; Comellas, G; Lemkau, L. R; Brea E. J; Berthold D. A; Zhou D; Rienstra, C. M. **2011.** (in prep) and leans heavily on unpublished ^{1}H-detection work from Dr. Donghua Zhou. "Accelerating structure determination of large microcrystalline and membrane proteins by proton detected solid-state NMR spectroscopy." Zhou, D. H; Berthold, D. A; Nieuwkoop, A, J; Sperling, L, J; Tang, M; Shah, G, J; Brea, E. J; Rienstra, C. M. **2011** (in prep).

We thank the National Institutes of Health (R01-GM073770 and ARRA supplement and NCRR Instruments Grant S10 RR025037-01) for financial support, and the School of Chemical Sciences NMR Facility at the University of Illinois.

4.2 Introduction

Amyloid fibrils are the pathological hallmark of many neurodegenerative diseases including Alzheimer's disease, Parkinson's disease and dementia with Lewy bodies. The structural study of these protein aggregates by X-ray and solution NMR is limited by the fact they are insoluble and do not form detracting crystals. The study of amyloid fibrils by SSNMR has a long history starting with Abeta by the Tycko lab[1, 2], transthyretin by the Griffin lab, human prion by Jaroniec[3-5], Het-S by Meier [6, 7] and alpha-synuclein (AS) by our group and others [8-12]

Proton detection combined with ^{1}H dilution and very-fast magic angle spinning (MAS) is an emerging area of SSNMR, which shows great promise for large gains in sensitivity versus heavy nuclei detection. Work in our lab as well as that done by the Reif group demonstrates that high quality ^{1}H detected spectra of nanocrystalline proteins can be obtained.[13, 14] Here we demonstrate that samples of uniformly $^{2}H^{13}C^{15}N$ labeled AS fibrils give similar resolution and sensitivity to those observed for nanocrystalline proteins.

4.3 Methods

4.3.1 Sample Preparation

For $^{2}H^{13}C^{15}N$ AS expression, Escherichia coli BL21(DE3) cells were first grown in LB medium, then inoculated into a $^{1}H_2O$-based M9 minimal medium supplemented with NH4Cl, glucose, BME vitamins, and yeast extract BioExpress (Cambridge Isotope Laboratory). The cells grew to a density of A600 of 1.65 and were diluted 10 fold in 100% $^{2}H_2O$ based medium of identical composition but containing $^{2}H,^{13}C,^{15}N$-BioExpress, $^{15}NH_4Cl$, and $^{2}H,^{13}C$-glucose. The cells again grew to a density of A600 of

1.7 and were diluted 10-fold into the same medium for a final concentration of ~99% ^2H_2O. When the cells reached a density of A600 of 1.16, expression was induced with 0.4 mM IPTG, and growth continued for 8 hours. The protein was purified according to reported protocol in 100% ^1H_2O buffer with a yield of 25 mg per 1 L of culture medium. Fibrils were prepared with seeding as previously reported[15] in buffers of 100% ^1H_2O and 25% ^1H_2O: 75% ^2H_2O.

4.3.2 ^{1}H Detected Pulse Sequences

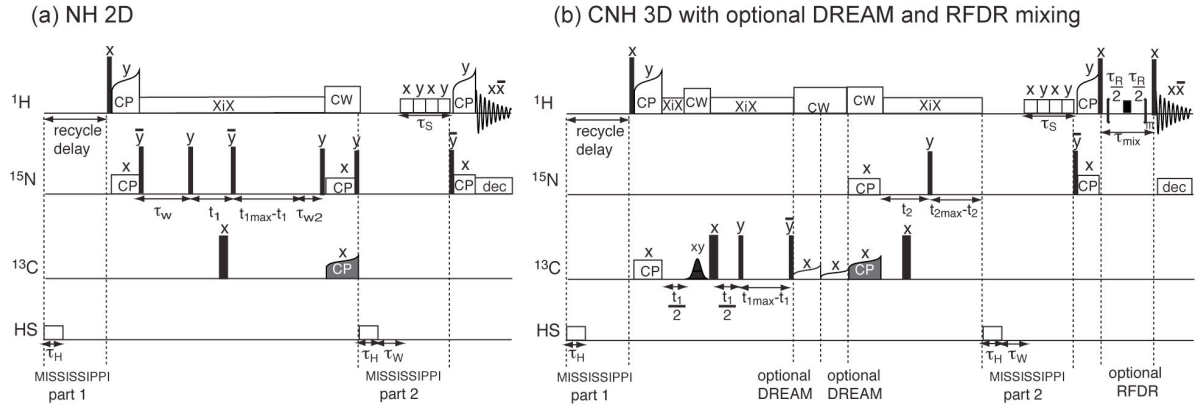

Figure 4.1 ^{1}H Detected SSNMR sequences for chemical shift assignment and distance determination.

(a) A NH 2D correlation sequence with MISSISSIPPI solvent suppression. (b) A CNH 3D correlation sequence with two optional C-C DREAM mixing periods and one optional proton-proton mixing period.

Pulse sequences for ^{1}H detected SSNMR spectroscopy developed by Donghua Zhou *et al.,*[13, 16, 17] are here expanded on to perform sequential chemical shift assignments (Donghua Zhou *et al.,* unpublished results). Figure 4.1a presents a NH 2D correlation sequence with MISSISSIPPI solvent suppression,[17] for use in accessing sample quality and resolution. Figure 4.1b presents an expanded 3D CNH sequence which can be used to determine ^{1}H-^{1}H correlations[16] through the optional RFDR

period,[18] and additionally provide CAcoNH, COcaNH and CBCANH 3D spectra for correlating backbone carbon chemical shifts to the amide proton-nitrogen pair through two option ^{13}C-^{13}C DREAM mixing periods.[19]

The NH 2D sequence (Figure 4.1a) begins with a homospoil pulse to suppress long T1 solvent signals, then proceeds to a proton to nitrogen cross polarization, followed by a constant time nitrogen evolution period with low power, XiX proton decoupling.[20, 21] A second homospoil pulse is followed by xy-4 presaturation pulses on proton to suppress solvent signals after which a nitrogen to proton cross polarization period produces a proton coherence for detection.

The CNH 3D sequence contains a similar solvent suppression regime with the addition of a carbon to nitrogen SPECIFIC CP[22] to provide the third carbon dimension. Adiabatic DREAM mixing[19] before the SPECIFIC-CP can provide CAcoNH or COcaNH type experiments to facilitate identification of CNH spin systems and CBCANH type experiments enable the observation of side chain carbon chemical shifts to assist in amino acid type identification.

4.4 ^{1}H detected SSNMR of AS with 100% Amide Protonation

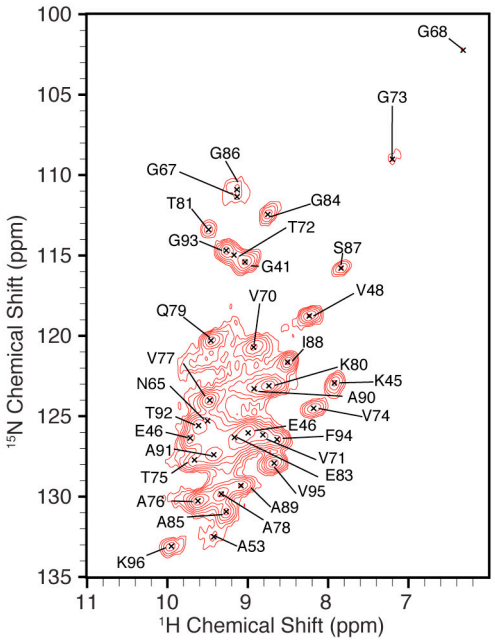

Figure 4.2. ^{15}N-^{1}H proton detected 2D spectrum of AS fibrils.

Spectrum acquired in 30 minutes with 19 ms of constant time ^{15}N evolution. Sample prepared with $^{2}H^{13}C^{15}N$ labeled AS monomer back exchanged during fibril formation with 100% $^{1}H_2O$ buffer.

^{1}H detected SSNMR experiments were performed on a 750 MHz Varian Inova spectrometer using a 1.6 mm FastMAS probe at MAS rates of 40 kHz. The CANH and CONH pulse sequences[13] with MISSISSIPPI[17] solvent suppression were employed. Samples of AS fibrils were prepared with uniform $^{2}H^{13}C^{15}N$ labeling and the amide ^{1}H sites back-exchanged with $^{1}H_2O$. Sample quality was first accessed through the use of ^{1}H detected ^{1}H-^{15}N 2D spectra (Figure 4.2). These spectra allow the ^{1}H linewidths and total signal intensity to be accessed before proceeding to three-dimensional experiments. This spectrum was acquired in 30 minutes on 400 nanomoles of AS, with 19 ms of constant time ^{15}N evolution and 50 ms of ^{1}H acquisition. ^{1}H linewidths for resolved peaks are 130 Hz, while ^{15}N linewidths are 50 Hz. While approximately 12

sites are resolved, it is obviously not possible to obtain chemical shift assignment from the 2D HN. In order to gain site-specific information for these systems it is necessary to go to higher dimensional experiments. Therefore, CNH 3Ds were acquired.

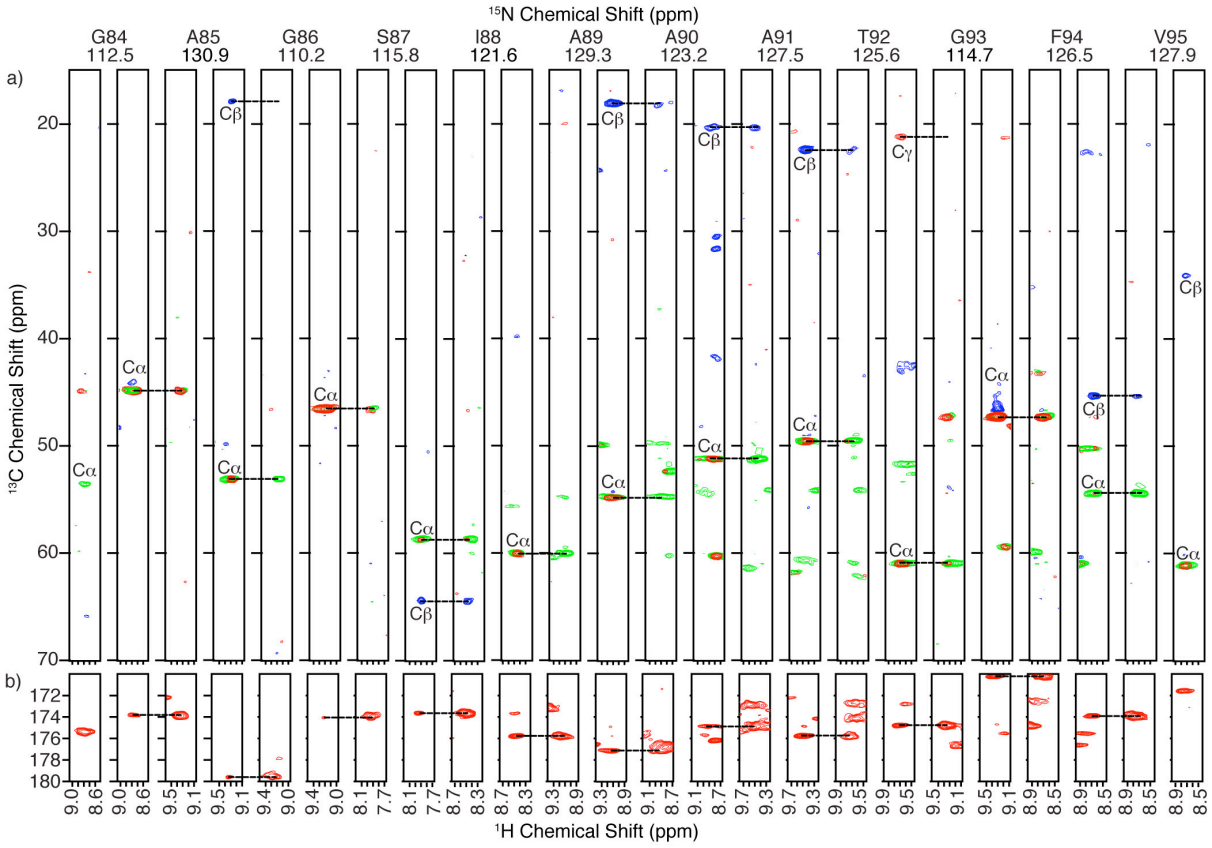

Figure 4.3. Backbone walk using six ^{1}H detected CNH 3D experiments.

Shown are the ^{1}H-^{13}C planes at the given ^{15}N frequency for each residue listed. The complementary CANH (green) and CBCANH (red positive, blue negative) data sets are overlaid and displayed next to the overlay of the CAcoNH (green) and CBCAcoNH (red positive, blue negative) spectra. Similarly the CONH and COcaNH (b) are also shown. Dotted lines aid the eye in making the correlation between the preceding (CAcoNH etc.) and current residue (CANH)s. CANH and CONH 3D's acquired in 18 hours each, CAcoNH in 38 hours, COcaNH in 25 hours, CBCANH in 42 hours and CBCAcoNH in 63 hours.

The CANH and CONH 3Ds are not sufficient to define backbone connectivity for the sites where the ^{15}N and ^{1}H pair is not unique. In order to assign backbone chemical shifts of AS a complete suite of ^{1}H detected experiments were acquired including

CAcoNH, COcaNH and CBCANH 3D's. These experiments contain one or two additional polarization transfers than those presented in earlier work. In addition to a ^{1}H-^{13}C CP period and a ^{13}C-^{15}N specific CP, ^{13}C-^{13}C DREAM mixing was used. The DREAM period is used for either CB-CA transfer, which provides side chain information to confirm amino acid type assignments, and CA-CO or CO-CA transfers, which enable the linking spin systems generated by CANH and CONH experiments. Using these experiments enabled the backbone walk for the core of AS fibrils residues 42 to 55 (except for H50 and G51) and residues 63 to 97 that include the new regions assigned with our ^{13}C-detected experiments.[23] The backbone walk from G84 to V95 is shown in Figure 4.3. In Figure 4.3a the CAcoNH (green) and CBCAcoNH (red positive, blue negative) are overlaid in the first (third etc.) column while the CANH (green) and CBCANH (red positive, blue negative) are overlaid in the second (forth etc). These ^{13}C-^{1}H planes are shown at the amide ^{15}N frequencies listed. In Figure 4.3b the CONH (first, third, etc) and COcaNH (second, forth, etc.) ^{13}C-^{1}H planes are shown for the same ^{15}N frequency. Matching carbon frequencies between one set of ^{15}N and ^{1}H frequencies in the intraresidue experiments (CANH, CBCANH, COcaNH) and another set of ^{15}N and ^{1}H frequencies in the interresidue experiments (CAcoNH, CBCAcoNH, and CONH) are indicated by the dotted lines. The combination of the CAcoNH and the CANH paired with the COcaNH and CONH provide connectivity along the backbone for all sites assigned.

However, these 4 experiments alone cannot provide site-specific assignments due to the difficulty in identifying any amino acid other than glycine by its CA and CO chemical shifts alone. Thus, the CBCANH and CBCAcoNH experiments that utilize ^{13}C-

^{13}C DREAM mixing[19] are needed. As can be seen in Figure 2, the CB of alanine residues, which appear as negative (blue) peaks around ~20 ppm, can be used to confidently identify those residues as Ala and Ser, respectively, allowing them to be placed in the primary sequence. Additionally, the CG2 of T92, appears as a positive (red) peak, in the same methyl region, confirming the residue as a Thr as two transfers are required to reach the CA. A full list of the backbone chemical shift assignments can be found in Table 4.1 at the end of the Chapter. These assignments agree with the ^{13}C-detection based assignments recently published by Gemma Comellas *et al*.[23]

4.5 ^{1}H-Detection of Samples with 25% Amide Protonation

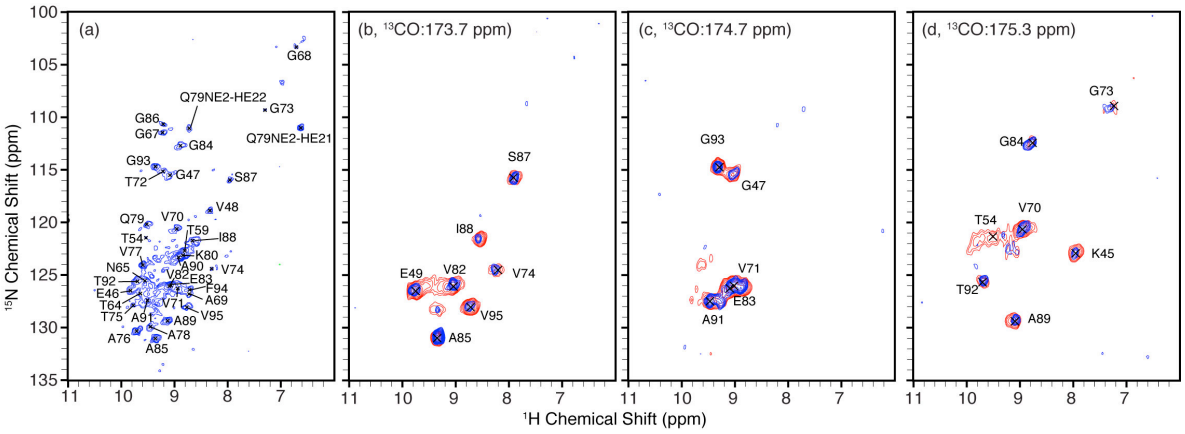

Figure 4.4. NH 2D and NH 2D planes from the CONH 3D.

NH 2D spectrum (a) acquired on the 25% ^{1}H diluted sample prepared with ^{2}H^{13}C^{15}N labeled AS monomer back exchanged with 25% $^{1}H_2O$ buffer. 3.4 hours of data with 19 ms of CT ^{15}N evolution. Planes from the CONH 3D (b-d) taken on the 25% ^{1}H diluted sample (blue) are overlaid with that taken on the 100% ^{1}H sample (red). Labels in (a) indicate amide ^{1}H assignments made from the 3D experiments while those in (b-d) indicate CO_{i-1}-N_i-HN_i assignments. CONH acquired in 46 hours while the CANH (not shown) was acquired in 38 hours.

For many regions of the protein the 100% $^{1}H_2O$ back-exchanged U-^{2}H^{13}C^{15}N fibrils provided sufficient resolution for a straightforward backbone walk. However, in some

regions the ~0.2 ppm ^{1}H linewidths resulted in overlap in the CNH 3D's. It has been shown that diluting the ^{1}H bath by exchanging the amide ^{1}H sites with less that 100% $^{1}H_2O$ can further improve ^{1}H linewidths.[24] Thus, in order to alleviate degeneracy fibrils were grown in buffer consisting of 25% $^{1}H_2O$ and 75% $^{2}H_2O$. This further decreased the number of ^{1}H-^{1}H couplings and resulted in narrower ^{1}H linewidths of 0.05 to 0.1 ppm for the NH 2D shown in Figure 4.4a. Although this sample had lower absolute intensity than the fibrils back exchanged with 100% $^{1}H_2O$, there was still sufficient sensitivity to acquire 3D CNH experiments. The benefit of the narrower ^{1}H linewidths is extremely apparent in the 3D CNH experiments. Figure 4.4b-d demonstrates ^{13}C planes from the CONH spectrum for the sample back exchanged with 100% $^{1}H_2O$ (red) and the sample back exchanged with 25% $^{1}H_2O$ (blue). The sharper ^{1}H linewidths in the 25% sample enable peaks like G93 and G47 (Figure 4.4c) to be easily distinguished. In addition, peaks like E49 and V82, which are broad in the spectrum taken on the 100% $^{1}H_2O$ sample, are better defined in the spectrum taken on the 25% sample, which reduces ambiguity in the ^{1}H chemical shift and increases confidence in assignments. This in turn significantly simplified the backbone walk procedures by reducing the number of possible pairings for each site. On the other hand T54, which is very weak in the spectrum taken on the 100% $^{1}H_2O$ sample (red, Figure 4.4d) does not appear in the spectrum taken on the 25% $^{1}H_2O$ sample. Additionally, the CNH 3D experiments with DREAM mixing are not sensitive enough to be practical on the 25% $^{1}H_2O$ sample. Thus both samples were needed to achieve all the assignments made.

4.6 Structural Restraints from ^{1}H Detected Spectra

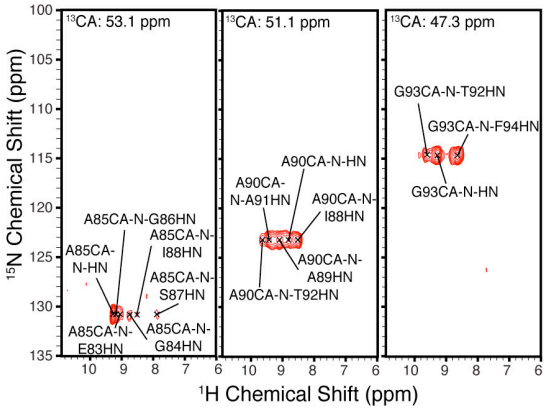

Figure 4.5. ^{1}H-^{15}N planes from a CANhH 3D with 8 ms of RFDR mixing.

Peaks labeled with potential assignments, F1 ^{1}H assignments are ambiguous. Data acquired in 38 hours on the 100% $^{1}H_2O$ sample.

In addition to being a powerful tool for chemical shift assignments of small quantities, ^{1}H detected SSNMR also benefits from the high gyromagnetic ratio of ^{1}H for long range ^{1}H-^{1}H mixing.[16] The ability to detect the ^{1}H after ^{1}H-^{1}H mixing, without the need for an additional cross polarization back to a heavy nuclei, further enhances the sensitivity gains when compared to CHHC type sequences which have been used so effectively in previous structural studies.[6, 25-27] 3D CNhH spectra, with between 2 and 8 ms of RFDR mixing were acquired. These spectra exhibit many cross peaks to each amide proton site in AS. Because ^{1}H linewidths in the 100% $^{1}H_2O$ sample are around 0.2 ppm there will be degeneracy in the final ^{1}H dimension. However the first proton, which is labeled by a ^{13}C and ^{15}N chemical shift, can be unambiguously identified, so the ambiguity is only in the third dimension. Figure 4.5 shows ^{15}N-^{1}H planes from such a CANhH 3D with 8 ms of RFDR mixing.[28] In this figure peaks are labeled with hypothetical ^{1}H assignments, however there are multiple possible assignments for the final ^{1}H

frequency. In particular for AS fibrils cross sheet ^{1}H-^{1}H distances would be expected to be shorter than i±2 distances, and many of the peaks assigned here to those correlations may well be long-range interactions. Modern protein structure calculations can utilize these types of ambiguous data to determine which correlations are being observed.[7, 25, 29] As such, we expect these data to provide valuable information for our ongoing structure determination efforts.

We have demonstrated that there is no inherent barrier to the application of ^{1}H-detected pulse sequences to amyloid fibrils that show similar linewidths to those observed in nanocrystalline protein samples. We show that 3D CNH experiments enable the assignment of backbone ^{13}C, ^{15}N and ^{1}HN chemical shifts. Dilution of the ^{1}H bath through back-exchanging samples with 25% $^{1}H_2O$ provided an increase in resolution by further sharpening ^{1}H linewidths, when compared to 100% $^{1}H_2O$ sample. Additionally, we also illustrate the detection of multiple ^{1}H-^{1}H distances in AS fibrils using ^{1}H detection with RFRD mixing.

Table 4.1 ^{1}H, ^{13}C and ^{15}N chemical shift assignments of U-$^{2}H^{13}C^{15}N$ - back exchanged AS fibrils

Residue	^{15}N	^{1}HN	$^{13}C'$	^{13}CA	^{13}CB	^{13}CG	$^{15}NE2$	$^{1}HE1/2$
S42	111.7	8.1	171.5	58.5	67.2			
K43	122.5	8.8	176.0	54.5	-			
T44	113.7	8.5	175.5	59.6	-			
K45	122.9	8.0	173.9	56.2	36.3			
E46	126.5	9.8	175.1	53.9	31.6			
G47	115.5	9.1	173.0	48.4				
V48	118.8	8.3	174.7	59.5	37.2			
V49	126.6	9.1	174.7	61.3	33.4			
V52	-	-	174.9	60.8	-			
A53	132.4	9.5	175.9	49.9	20.9			
T54	121.4	9.5	-	61.6	-			
V55	126.5	9.5	174.3	61.0	-			
T59	122.6	8.8	-	67.0	-			
V63	-		175.2	61.0	-			
T64	126.5	9.6	172.7	62.2	-			
N65	125.3	9.6	172.9	51.7	42.7			
V66	127.2	9.5	178.1	60.7	-			
G67	111.3	9.2	172.9	46.4				
G68	102.4	6.5	172.6	43.2				
A69	126.7	8.8	175.5	50.2	22.8			
V70	120.7	8.9	174.8	59.9	35.1			

Table 4.1, continued

Residue	^{15}N	^{1}HN	^{13}C'	^{13}CA	^{13}CB	^{13}CG	^{15}NE2	^{1}HE1/2
V71	126.2	8.9	176.7	61.0	34.6			
T72	115.1	9.2	175.6	59.4	-	21.3		
G73	109.2	7.3	173.8	43.8				
V74	124.5	8.2	175.4	61.3	34.1			
T75	127.8	9.7	172.2	61.8	-	20.8		
A76	130.3	9.7	174.5	49.7	20.8			
V77	124.0	9.5	173.2	60.5	35.2			
A78	129.9	9.4	176.5	49.9	24.3			
Q79	120.3	9.5	176.8	52.4	32.5		111.0	6.6/8.7
K80	123.1	8.8	176.2	60.3	31.6			
T81	113.5	9.5	173.8	61.0	-	21.9		
V82	126.0	9.0	175.0	61.4	33.2			
E83	125.9	9.0	175.4	53.6	33.2			
G84	112.5	8.8	173.8	44.8				
A85	130.9	9.3	179.5	53.1	17.9			
G86	110.8	9.2	174.0	46.5				
S87	115.9	7.9	173.7	58.7	64.4			
I88	121.6	8.5	175.8	60.0	39.8			
A89	129.3	9.1	177.2	54.8	18.1			
A90	123.2	8.9	174.9	51.2	20.4			
A91	127.4	9.5	175.8	49.5	22.4			

Table 4.1, continued

Residue	^{15}N	^{1}HN	^{13}C'	^{13}CA	^{13}CB	^{13}CG	^{15}NE2	^{1}HE1/2
T92	125.6	9.7	174.8	61.0	-	21.3		
G93	114.7	9.3	170.2	47.3				
F94	126.5	8.7	173.9	54.4	45.4			
V95	127.9	8.7	171.5	61.1	34.1			
K96	133.1	9.9	173.2	54.9	-			
K97	130.2	9.9	175.6	54.0	-			

4.7 References

1. Petkova, A. T.; Leapman, R. D.; Guo, Z. H.; Yau, W. M.; Mattson, M. P.; Tycko, R., "Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils." *Science* 2005, *307* (5707), 262-265.
2. Paravastu, A. K.; Qahwash, I.; Leapman, R. D.; Meredith, S. C.; Tycko, R., "Seeded growth of beta-amyloid fibrils from Alzheimer's brain-derived fibrils produces a distinct fibril structure." *Proc. Natl. Acad. Sci. U. S. A.* 2009, *106* (18), 7443-7448.
3. Jaroniec, C. P.; MacPhee, C. E.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G., "High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy." *Proc. Natl. Acad. Sci. U. S. A.* 2004, *101* (3), 711-716.
4. Helmus, J. J.; Surewicz, K.; Nadaud, P. S.; Surewicz, W. K.; Jaroniec, C. P., "Molecular conformation and dynamics of the Y145Stop variant of human prion protein." *Proc. Natl. Acad. Sci. U. S. A.* 2008, *105* (17), 6284-6289.
5. Helmus, J. J.; Surewicz, K.; Surewicz, W. K.; Jaroniec, C. P., "Conformational Flexibility of Y145Stop Human Prion Protein Amyloid Fibrils Probed by Solid-State Nuclear Magnetic Resonance Spectroscopy." *J. Am. Chem. Soc.* 2010, *132* (7), 2393-2403.
6. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H., "Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core." *Science* 2008, *319* (5869), 1523-1526.
7. Van Melckebeke, H.; Wasmer, C.; Lange, A.; Ab, E.; Loquet, A.; Bockmann, A.; Meier, B. H., "Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state NMR spectroscopy." *J. Am. Chem. Soc.* 2010, *132* (39), 13765-75.
8. Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M., "Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR." *Proc. Natl. Acad. Sci. USA* 2005, *102* (44), 15871-15876.

9. Kloepper, K. D.; Hartman, K. L.; Ladror, D. T.; Rienstra, C. M., "Solid-state NMR spectroscopy reveals that water is nonessential to the core structure of alpha-synuclein fibrils." *J. Phys. Chem. B* 2007, *111* (47), 13353-13356.
10. Kloepper, K. D.; Zhou, D. H.; Li, Y.; Winter, K. A.; George, J. M.; Rienstra, C. M., "Temperature-dependent sensitivity enhancement of solid-state NMR spectra of alpha-synuclein fibrils." *J. Biomol. NMR* 2007, *39* (3), 197-211.
11. Vilar, M.; Chou, H.-T.; Luhrs, T.; Samir, M.; Riek-Loher, D.; Verel, R.; Manning, G.; Stahlberg, H.; Riek, R., "The fold of alpha-synuclein fibrils." *Proc. Natl. Acad. Sci. U. S. A.* 2008, *105* (25), 8637-8642.
12. Loquet, A.; Giller, K.; Becker, S.; Lange, A., "Supramolecular Interactions Probed by C-13-C-13 Solid-State NMR Spectroscopy." *J. Am. Chem. Soc.* 2010, *132* (43), 15164-15166.
13. Zhou, D. H.; Shah, G.; Cormos, M.; Mullen, C.; Sandoz, D.; Rienstra, C. M., "Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning." *J. Am. Chem. Soc.* 2007, *129* (38), 11791-11801.
14. Linser, R.; Fink, U.; Reif, B., "Proton-detected scalar coupling based assignment strategies in MAS solid-state NMR spectroscopy applied to perdeuterated proteins." *J. Magn. Reson.* 2008, *193* (1), 89-93.
15. Kloepper, K. D.; Woods, W. S.; Winter, K. A.; George, J. M.; Rienstra, C. M., "Preparation of alpha-synuclein fibrils for solid-state NMR: Expression, purification, and incubation of wild-type and mutant forms." *Protein Expres. Purif.* 2006, *48* (1), 112-117.
16. Zhou, D. H.; Shea, J. J.; Nieuwkoop, A. J.; Franks, W. T.; Wylie, B. J.; Mullen, C.; Sandoz, D.; Rienstra, C. M., "Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy." *Angew. Chem., Int. Ed.* 2007, *46* (44), 8380-8383.
17. Zhou, D. H.; Rienstra, C. M., "High-performance solvent suppression for proton detected solid-state NMR." *J. Magn. Reson.* 2008, *192* (1), 167-172.
18. Bennett, A. E.; Rienstra, C. M.; Griffiths, J. M.; Zhen, W. G.; Lansbury, P. T.; Griffin, R. G., "Homonuclear radio frequency-driven recoupling in rotating solids." *J. Chem. Phys.* 1998, *108* (22), 9463-9479.
19. Verel, R.; Ernst, M.; Meier, B. H., "Adiabatic dipolar recoupling in solid-state NMR: the DREAM scheme." *J. Magn. Reson.* 2001, *150* (1), 81-99.
20. Detken, A.; Hardy, E. H.; Ernst, M.; Meier, B. H., "Simple and efficient decoupling in magic-angle spinning solid- state NMR: the XiX scheme." *Chem. Phys. Lett.* 2002, *356* (3-4), 298-304.
21. Ernst, M.; Samoson, A.; Meier, B. H., "Low-power XiX decoupling in MAS NMR experiments." *J. Magn. Reson.* 2003, *163* (2), 332-339.
22. Baldus, M.; Geurts, D. G.; Hediger, S.; Meier, B. H., "Efficient N-15-C-13 polarization transfer by adiabatic-passage Hartmann-Hahn cross polarization." *J. Magn. Reson. A* 1996, *118* (1), 140-144.
23. Comellas, G.; Lemkau, L. R.; Nieuwkoop, A. J.; Kloepper, K. D.; Ladror, D. T.; Ebisu, R.; Woods, W. S.; Lipton, A. S.; George, J. M.; Rienstra, C. M., "Structured Regions of alpha-Synuclein Fibrils Include the Early-Onset Parkinson's Disease Mutation Sites." *J. Mol. Biol.* 2011.
24. Akbey, U.; Lange, S.; Franks, W. T.; Linser, R.; Rehbein, K.; Diehl, A.; van Rossum, B. J.; Reif, B.; Oschkinat, H., "Optimum levels of exchangeable protons in

perdeuterated proteins for proton detection in MAS solid-state NMR spectroscopy." *J. Biomol. NMR* 2010, *46* (1), 67-73.
25. Loquet, A.; Bardiaux, B.; Gardiennet, C.; Blanchet, C.; Baldus, M.; Nilges, M.; Malliavin, T.; Bockmann, A., "3D structure determination of the Crh protein from highly ambiguous solid-state NMR restraints." *J. Am. Chem. Soc.* 2008, *130* (11), 3579-3589.
26. Franks, W. T.; Wylie, B. J.; Schmidt, H. L. F.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M., "Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR." *Proc. Natl. Acad. Sci. U. S. A.* 2008, *105* (12), 4621-4626.
27. Jehle, S.; Rajagopal, P.; Bardiaux, B.; Markovic, S.; Kuhne, R.; Stout, J. R.; Higman, V. A.; Klevit, R. E.; van Rossum, B. J.; Oschkinat, H., "Solid-state NMR and SAXS studies provide a structural basis for the activation of alpha B-crystallin oligomers." *Nat. Struct. Mol. Biol.* 2010, *17* (9), 1037-U1.
28. Bennett, A. E.; Ok, J. H.; Griffin, R. G.; Vega, S., "Chemical shift correlation spectroscopy in rotating solids: Radio-frequency dipolar recoupling and longitudinal exchange." *J. Chem. Phys.* 1992, *96*, 8624-8627.
29. Nieuwkoop, A. J.; Rienstra, C. M., "Supramolecular Protein Structure Determination by Site-Specific Long-Range Intermolecular Solid State NMR Spectroscopy." *J. Am. Chem. Soc.* 2010, *132* (22), 7570-+.

Chapter 5

Progress Towards the High-Resolution Structure of Wild-type Human Alpha-Synuclein Fibrils

5.1 Notes and Acknowledgments

Portions of this chapter are reprinted with permission from a jointly authored publication, "Structured Regions of alpha-Synuclein Fibrils Include the Early-Onset Parkinson's Disease Mutation Sites", Comellas, G.; Lemkau, L. R.; Nieuwkoop, A. J.; Kloepper, K. D.; Ladror, D. T.; Ebisu, R.; Woods, W. S.; Lipton, A. S.; George, J. M.; Rienstra, C. M., *J. Mol. Biol.* **2011**, *411,* 881-895. Copyright 2011, Elsevier. These results build upon studies of past and present group members to obtain confident site-specific ^{13}C and ^{15}N chemical shift assignments from Kloepper and Comellas.[1, 2] Sample preparation efforts were supported by Dr. Deborah Berthold, Luisel Lemkau and several undergraduate researchers, Kevin Hartman, Daniel Ladror, Reika Ebisu, Shin Lee Wook and Lars Rikardsen, and complementary studies of Lemkau et al. to assign mutant forms (A30P, E46K, and A53T) of alpha-synuclein (AS).

This work was supported by the National Institutes of Health (R01-GM073770, R01-GM073770 ARRA supplement and NCRR Instruments Grant S10 RR025037-01). Gemma Comellas was a Caja Madrid Foundation Graduate Fellow. A portion of the research was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National

Laboratory with the assistance of Andrew Lipton, Sarah Burton and David Hoyt. Electron micrographs were carried out in the Frederick Seitz Materials Research Laboratory Central Facilities, University of Illinois, which are partially supported by the U.S. Department of Energy under grants sDE-FG02-07ER46453 and DE-FG02-07ER46471. NMR experiments were performed at the School of Chemical Sciences NMR Facility at the University of Illinois, with the assistance of Jen Rapp.

5.2 Introduction

The protein alpha-synuclein (AS) is the primary proteinaceous component of Lewy bodies, which are the pathological hallmark of Parkinson's disease (PD).[3] Several point mutations of the AS gene are found to accelerate the onset of the disease in families with early onset PD.[4-7] Nevertheless, patients with mutant forms of AS account for only a very small fraction of PD cases.[8] Thus, the connections between AS and PD are not fully understood, and relatively little is known definitively about the functional role AS plays in normal dopaminergic neuron function, as well as the specific dysfunction of AS that initiates neurodegeneration. It even remains in debate whether AS fibrils are the cause or result of PD. Recent studies have indicated that cell-to-cell transmission of AS aggregates may play a role in the progression of PD,[9] and that exogenous AS fibrils can seed the formation of Lewy-type inclusions in cultured cells.[10]

To solve the structure of AS fibrils at atomic resolution would provide useful information in unraveling this mystery. AS fibrils are insoluble and do not form diffraction quality crystals, so the only route to atomic-resolution structural information is SSNMR. We have focused efforts on the wild-type (wt) human form of AS, and plan to examine changes in structure caused by mutations in order to understand the role these sites play structurally.

As discussed in Chapter 1, much progress has been made in the field of fibril structure determination in the last few years. One example of a major success is the structure of the HET-s prion from Meier and co-workers.[11] Notably, this 72-residue protein fragment is half the size of AS. To solve the structure of HET-s required 437 experimental restraints for the ~40 structured residues, building from previous

assignment and secondary structure analysis from the same group.[12] These studies illustrated the progression from low resolution model to full structure. The model was based only on secondary chemical shifts and signal intensities and reported the β-solenoid motif but lacked information about intermolecular arrangements and did not define the inter-sheet packing arrangements, side-chain conformations, and hydrogen bonding patterns. Unambiguous intramolecular and intermolecular distance information proved critical for defining the HET-s structure and subsequently refining it to atomic resolution.[13-15] Here I will present low-resolution models of AS, generated on the basis of secondary structure data (chemical shifts), and substantial quantities of SSNMR data acquired to restrain ^{13}C-^{13}C, ^{15}N-^{13}C and ^{1}H-^{1}H distances. The distance information then is utilized to test the models. Unique agreement between the structural models and distance restraints will lead to an atomic resolution structure of AS fibrils.

5.3 Structural Information from Other Sources

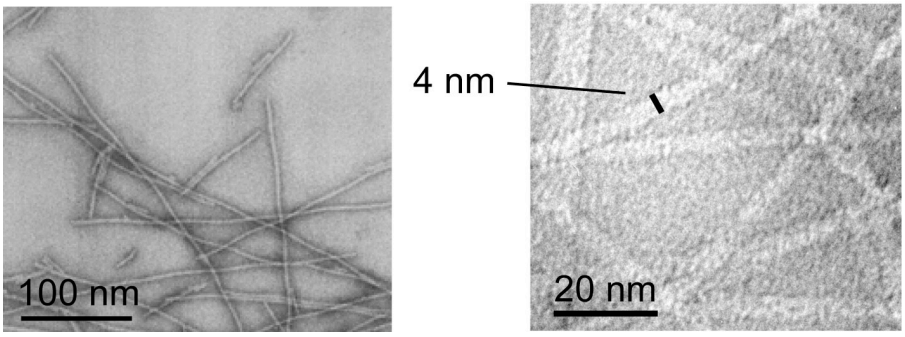

Figure 5.1: Transmission electron micrographs of wt AS fibrils.

A second scale bar indicates the 4 nm minimum width observed for these fibrils. EM data acquired by Gemma Comellas at the Frederick Seitz Materials Research Laboratory Central Facilities, University of Illinois at Urbana-Champaign. Figure reproduced with permission from Comellas *et al*.[2]

In order to frame a discussion of the atomic resolution structure of AS fibrils, it is important to consider the lower resolution information that is already available. Figure 5.1 shows electron micrographs (EM) of wt AS fibrils fibrilized in a manner identical to SSNMR preparations,[16] acquired by Gemma Comellas.[2] The micrographs indicate that single strands of wt AS fibrils have a fibril width of about 4 to 4.5 nm and demonstrate a high degree of homogeneity and long-range order. UV absorption spectroscopy provides another low-resolution structural restraint; the absorption maximum of the backbone C=O band is sensitive to the hydrogen bonding environment of parallel or anti-parallel beta-sheets, and for wt AS fibrils these results support a composition of mostly parallel beta-sheets.[17] Combining these results with the geometrical properties of beta-sheet peptides (a residue in beta-sheet secondary structure extends about 3.5 Å

along the peptide chain), we expect a beta-sheet of length between 11 and 13 residues, providing an important constraint for model building.

One major challenge that arises in performing structural studies of any protein, and especially protein fibrils, is obtaining consistent batch-to-batch morphology. It is likely that there are several forms of AS fibrils, perhaps more than the number of groups studying them.[18-21] Two other groups have presented chemical shift assignments,[21, 22] which differ substantially from those determined by our group (Comellas *et al.*,[2]). We attribute this to differences in the details of the fold depending on salt, temperature, incubation method and whether samples are seeded. Previous structural studies were not at sufficiently high resolution to identify these subtle differences, e.g., using site directed spin labeling[23] or a combination of SSNMR and HD exchange data.[20] These studies are consistent with the idea that residues ~30 to 100 form the structured core of the fibrils, but are not sufficiently high resolution to be compared directly with our outcomes. Given the high degree of structural polymorphism observed in the AS fibrils of others,[22] and fibrils in general[24]–due to factors such as buffer composition and agitation–and the fact that no other group utilizes the protocol developed by Kloepper to specifically avoid such sample to sample variation,[16] we must consider the structure of wt AS fibrils investigated here separately from the data presented by others.

5.4 Initial Models of AS Fibrils

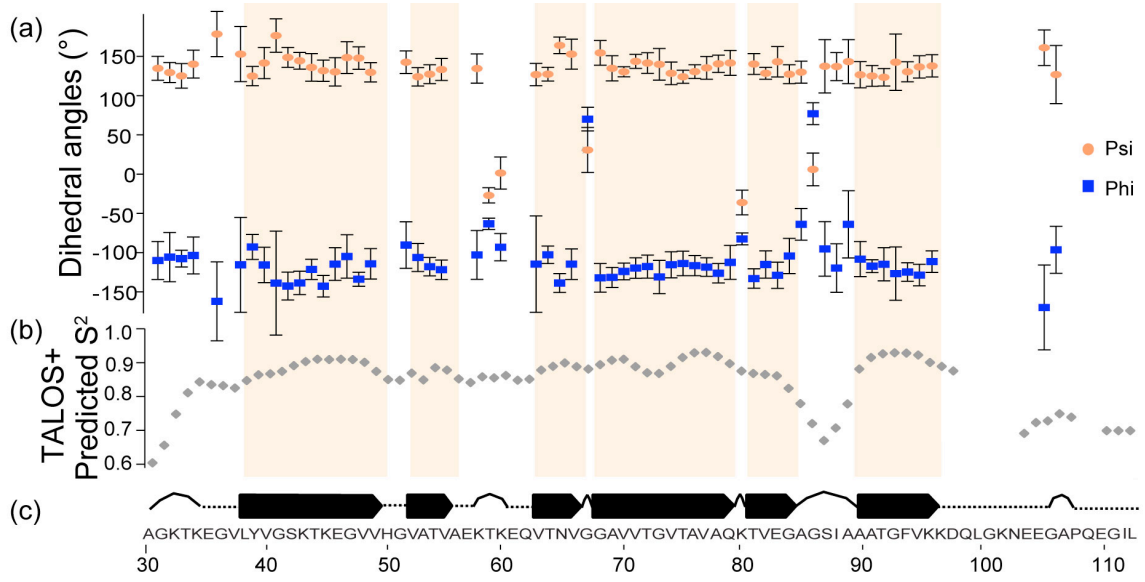

Figure 5.2: TALOS+ results from the chemical shifts of wild type human AS.

Site specific chemical shifts[2] were used to predict a) backbone dihedral angles, b) order parameters and c) secondary structure (sheet thick arrows, loops solid line, no prediction dotted lines) using the program TALOS+.[25] Figure reproduced with permission from Comellas *et al*.[2]

Semi-empirical relationships between secondary chemical shifts and secondary structure (via the program TALOS+[25]) allowed us to construct a model consisting of the location of beta-strands with respect to the sequence (Figure 5.2). The TALOS+ results indicate well-structured regions in AS fibrils from residues 31 to 100, with several flexible loops. A motif of 12-residue strands, connected by a short loop to a ~5-residue strand, is repeated (once between residues 37 and 55 and again from residues 67 to 85).[2] The 12-residue strand length is consistent with the 11- to 13-residue fibril width determined by EM and UV data, and therefore it factors heavily in early models of the fibril structure.

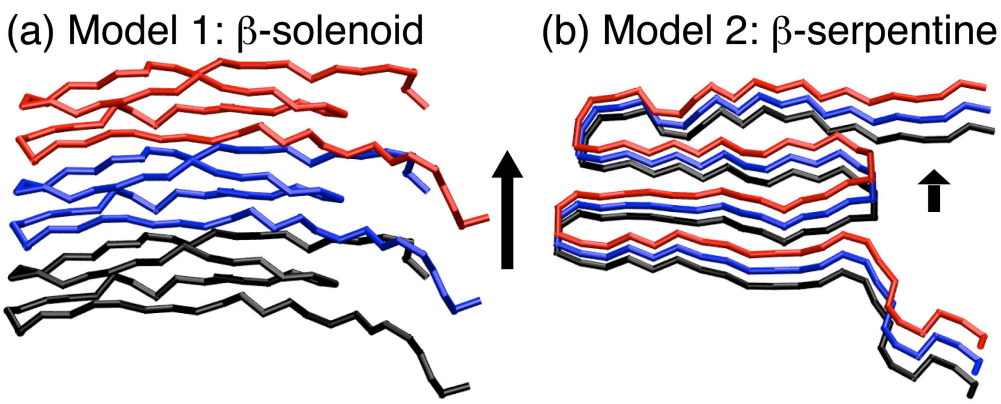

Figure 5.3: Models of the structure of wt AS fibrils.

Two models of the AS fibril structure one (a) based on a beta-solenoid and the other (b) based on a beta-serpentine fold. Three monomers (residues 29 to 100) of AS are shown for each. The beta-solenoid model is show with the axis of fibril propagation, indicated by the arrow, running parallel to the page, while the beta-serpentine model is shown with the axis of propagation coming out of the page.

EM, UV and TALOS+ data together are not sufficient at this stage to differentiate between the two most likely fibril topologies. This is illustrated in Figure 5.3 where two AS fibril models consistent with the TALOS results and a 4.5 nm fibril width consisting of parallel beta-sheets are shown. They differ primarily in that Figure 5.3a is a beta-solenoid and Figure 5.3b is a beta-serpentine structure. In a beta-solenoid, each AS monomer forms two layers of the fibril by forming an intramolecular helical arrangement (a beta-helix), with both inter- and intra-molecular beta-sheets with an i to i+30±1 registry (e.g., residue 37 is near residue 67). In the beta-serpentine, all beta-sheets are formed of strands from different molecules (i.e., an intermolecular beta-sheet) and are parallel and in register. Standard EM measurements are unable to distinguish between these two models, but using more precise scanning tunneling EM (STEM) methods may provide the mass per length (MPL); such studies are ongoing (Comellas et al., unpublished).

5.5 Survey of Experimental SSNMR Structural Data

We proceeded to collect a variety of internuclear distance information with 2D and 3D SSNMR experiments (Table 5.1). Experiments performed with undiluted samples yield both intra- and inter-molecular distance restraints, and provide the highest experimental sensitivity. Experiments performed with diluted samples focus primarily on the intramolecular distances, and experiments performed with mixtures of differently labeled samples highlight intermolecular contacts. Intermolecular SSNMR experiments would show either characteristic i to i+/-29/30 or i to i+/-1 cross peaks.

Table 5.1: SSNMR experiments for distance restraints

Experiment	Number of Dims.	Mixing Scheme	Labeling Pattern#	Mixing Time	Acq. Time (hours)	Field (MHz)	Distance Type
CC	2D	DARR[26]	U^{13}C,^{15}N dil.	400 ms	93	750	Intra
CC	2D	DARR	2-gly dil.	400 ms	144	750	Intra
CC	2D	DARR	2-gly	400 ms	78	750	Both
CC	2D	DARR	2-gly	500 ms	56	900	Both
CC	2D	DARR	1,3-gly	500 ms	42	900	Both
CC	2D	DARR	U^{13}C,^{15}N	250 ms	34	900	Both
CC	2D	DARR	1,3-gly	500 ms	107	750	Both
CC	2D	DARR	1,3-gly	250 ms	107	750	Both
NCACX	3D	DARR	U^{13}C,^{15}N	400 ms	280	750	Both
NCACX	3D	DARR	2-gly	500 ms	115	600	Both
NCOCX	3D	DARR	2-gly	250 ms	77	600	Both
NCOCX	3D	DARR	1,3-gly	250 ms	134	600	Both
CC	2D	ChhC[27]	U^{13}C,^{15}N	150 us	172	600	Both
CC	2D	ChhC	U^{13}C,^{15}N	300 us	71	600	Both
CC	2D	ChhC	U^{13}C,^{15}N	450 us	197	600	Both
CC	2D	ChhC	U^{13}C,^{15}N dil.	180 us	240	750	Intra
CC	2D	ChhC	U^{13}C,^{15}N	180 us	106	750	Both
NC	2D	NhhC[27]	U^{13}C,^{15}N	180 us	196	500	Both
NC	2D	NhhC	U^{13}C,^{15}N dil.	180 us	194	500	Intra
NC	2D	NhhC	Mixed 2-gly	180 us	112	500	Inter
NC	2D	TEDOR[28]	1,3-gly	1.44 to 14.4 ms	145	500	Both
NC	2D	TEDOR	2-gly	1.44 to 14.4 ms	198	500	Both
NC	2D	TEDOR	Mixed 1,3-gly	5.66 to 14.4 ms	154	500	Inter
NC	2D	TEDOR	Mixed 2-gly	14.4 ms	120	500	Inter

Table 5.1, continued

Experiment	Number of Dims.	Mixing Scheme	Labeling Pattern[#]	Mixing Time	Acq. Time (hours)	Field (MHz)	Distance Type
CANhH	3D	RFDR[29]	CDN	8 ms	38	750	Both
CANhH	3D	RFDR	CDN	6 ms	36	750	Both
CONhH	3D	RFDR	CDN	6 ms	18	750	Both
NhH	2D	RFDR	CDN	6 ms	7	750	Both

[#]Labeling Pattern Key: U^{13}C,^{15}N: Uniformly ^{13}C, ^{15}N; U^{13}C,^{15}N dil. : Uniformly ^{13}C, ^{15}N diluted 1:3 in n.a.; 1,3-gly: 1,3-^{13}C-glycerol, ^{15}N; 2-gly: 2-^{13}C-glycerol, ^{15}N; Mixed 1,3-gly : 50:50 1,3-^{13}C-glycerol, n.a. nitrogen: n.a. carbon, ^{15}N ; Mixed 2-gly: 50:50 2-^{13}C-glycerol, n.a. nitrogen: n.a. carbon, ^{15}N ; CDN: ^{13}C^{2}H^{15}N labeled back exchanged with 100% $^{1}H_2O$.

In order to determine the structure of AS fibrils to more precision than the models just presented, large amounts of SSNMR data will be needed. Table 5.1 presents a list of the several months of SSNMR experiments acquired on wt AS fibrils for the purpose of determining distance restraints, in collaboration with Gemma Comellas, using 500, 600, 750 MHz spectrometers at UIUC as well as the 900 MHz spectrometer at Pacific Northwest National Laboratories in Richland, WA. As noted above, some samples were uniformly enriched in ^{13}C, ^{15}N and/or ^{2}H and undiluted. A second set was uniformly enriched and diluted in natural abundance protein. A third set of samples was isotopically mixed, typically with ^{15}N enriched protein was mixed 50:50 with ^{13}C enriched protein, to yield primarily intermolecular correlations. As discussed by Van Melckebeke *et al.,*[15] while a majority of the correlations used in a fibril structure calculation can come from uniformly enriched samples, without diluted and mixed samples it will be impossible to determine inter- from intramolecular contacts. In addition, while the higher sensitivity of 2D spectra give the highest probability of observing long-range correlations, the added resolution provided by 3D experiments is also necessary to ensure a sufficient number of unambiguous correlations.

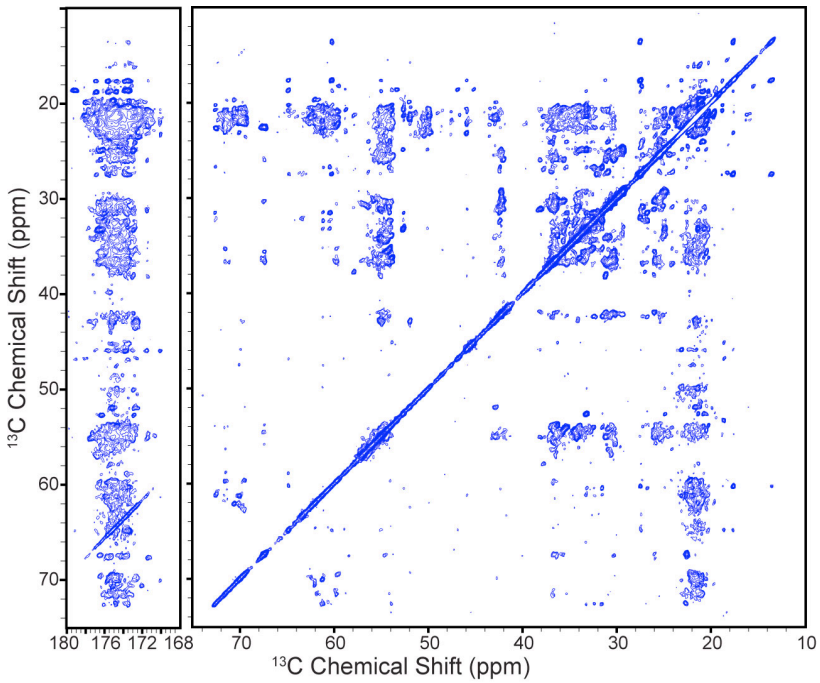

Figure 5.4: Carbon-carbon 2D spectrum of 1,3-^{13}C-glycerol labeled wt AS.

Spectrum acquired at 900 MHz ^{1}H-frequency with 500 ms of DARR mixing.

Among the classes of experiments listed in Table 5.1, the most numerous are the carbon-carbon 2D spectra. As discussed in Chapter 1, this homonuclear correlation experiment is perhaps the highest sensitivity and most robust method for detecting inter-residue correlations. When this experiment is applied to samples with the 1,3-^{13}C and 2-^{13}C-glycerol labeling scheme, spectral sensitivity and resolution are further enhanced (but elimination of most one-bond ^{13}C-^{13}C scalar couplings) and provide a preponderance of interresidue and long-range distance restraints.[13] Figure 5.4 shows a carbon-carbon 2D with 500 ms of DARR mixing on a sample of 1,3-^{13}C-glycerol labeled AS acquired at 900 MHz at Pacific Northwest National Laboratories. Individual peak linewidths are ~0.3 ppm, due to the glycerol labeling and high magnetic field. Furthermore, due to the long ^{13}C-^{13}C mixing period of 500 ms, many correlations can be

observed to each carbon site. For example, at F1=18.5 ppm the Ala89 CB site shows correlations to at least 6 carbonyl and at least a dozen aliphatic carbons in the F2 dimension. However, in regions of the spectrum where multiple carbons are overlapped, for example between 20 and 24 ppm, individual peaks are rarely resolved. Despite this limitation, carbon-carbon 2D spectra of this and other samples provide a large number of distance restraints for structure calculations.

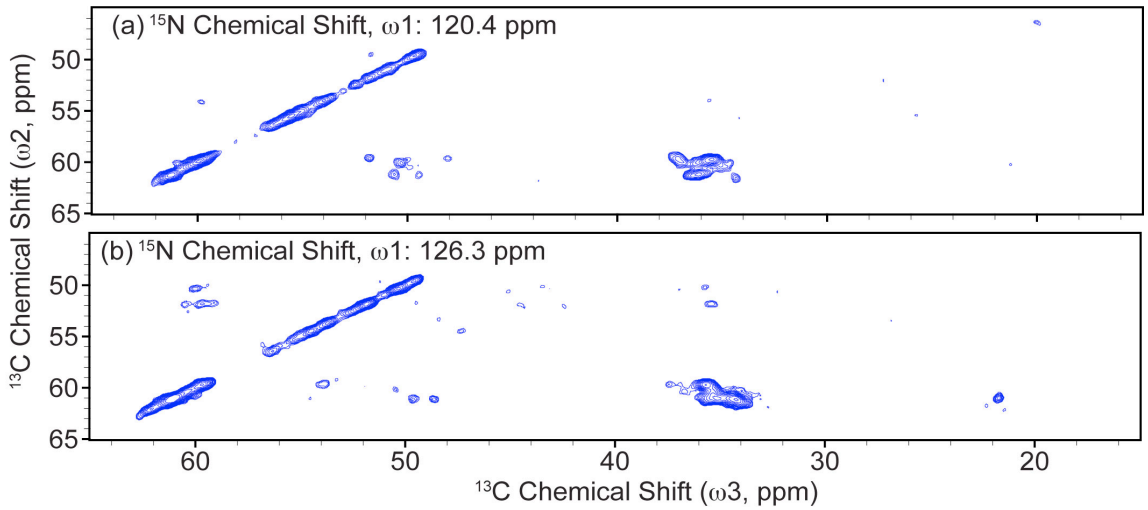

Figure 5.5: Carbon-carbon planes from a 3D CANCX spectrum.

2D planes at the ^{15}N frequencies of (a) 120.4 and (b) 126.3 ppm, from a CANCX 3D spectrum with 500ms of DARR mixing acquired on a 600 MHz ^{1}H-frequency spectrometer with a uniformly ^{13}C,^{15}N labeled sample.

We also performed 3D versions of this experiment, using a ^{15}N dimension to increase the percentage of signals with uniquely resolved long-range correlations. Through the use of SPECFIC-CP[30] to obtain selective N-CA or N-C' transfers, a NCACX or NCOCX type experiment can be performed. While utilized extensively at shorter mixing times for sequential backbone assignments,[31] in this context the ^{15}N dimension is utilized to provide resolution of carbon-carbon correlations. For example, in Figure 5.5

the peaks clustered around 61 ppm (ω2) and 50 ppm (ω3) would be overlapped in a carbon-carbon 2D. However, by frequency labeling the ^{15}N frequencies they are resolved. In this manner, many long-range carbon-carbon correlations can be resolved despite the relative lack of ^{13}C chemical shift dispersion in AS.

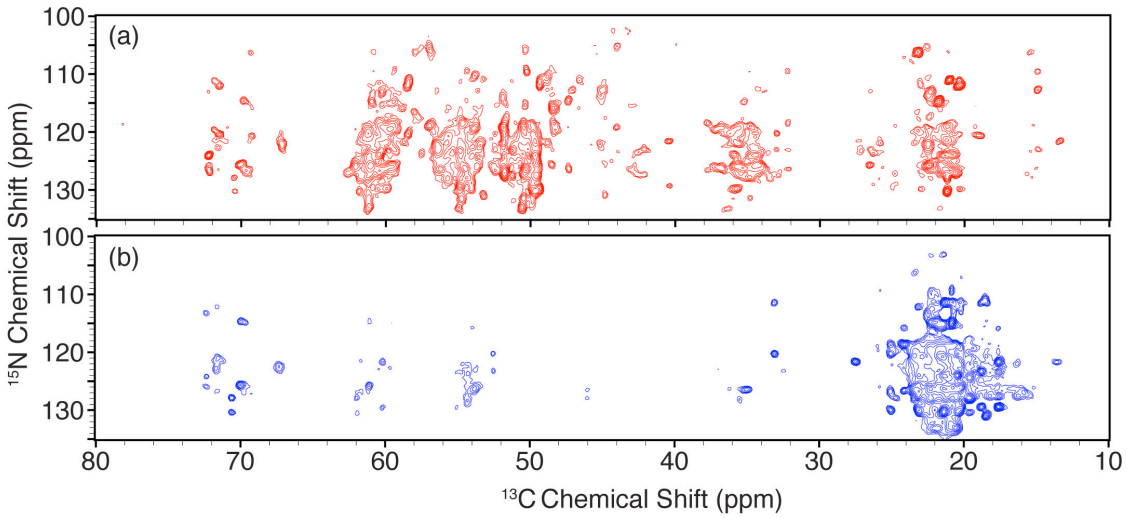

Figure 5.6: 2D ZF-TEDOR plane of glycerol labeled AS fibrils.

TEDOR planes acquired with 14.4 ms of ^{15}N-^{13}C mixing at 500 MHz on (a) 2-^{13}C-glycerol, uniform ^{15}N labeled and (b) 1,3-^{13}C-glycerol, uniform ^{15}N labeled AS samples.

The 3D ZF-TEDOR pulse sequence can also provide long-range distance information as discussed in depth in Chapter 2. 2D TEDOR planes were acquired on glycerol labeled wt AS samples. Figure 5.6a shows the spectrum of a 2-^{13}C-glycerol labeled sample while Figure 5.6b shows the spectrum of a 1,3-^{13}C-glycerol labeled sample. Both spectra were acquired with 14.4 ms of ^{15}N-^{13}C mixing, the longest TEDOR mixing time acquired on these samples. In comparison to the TEDOR spectra of GB1 (Chapter 2, Figure 2.5) at this mixing time there are significantly fewer correlations observed in AS. While part of this is due to the lower signal to noise ratio of these AS spectra, the topology of the AS fibrils is likely also a factor. The sensitivity of these

spectra is not sufficient to see the ~8 Å distances observed in GB1, but 4 to 5 Å distances should be observable. We attribute the smaller number of long-range correlations here to the parallel in registry structure that is likely to be present; in this geometry, the closest amino acid residue in the neighboring molecule is the same as that present at the originating residue. Thus, the source and destination residues in the polarization transfer event have identical chemical shifts. The fact that the cross-sheet partners have the same chemical shifts as the inter-residue and sequential peaks observed at short TEDOR mixing time indicates a parallel, in-register intermolecular arraignment of AS monomers.

5.6 Test of the Intermolecular Parallel Registry

To test whether AS fibrils are in a parallel registry, we performed intermolecular TEDOR experiments as discussed in Chapter 3. Experiments were performed using several isotopic labeling patterns, following the logic introduced in the previous section.

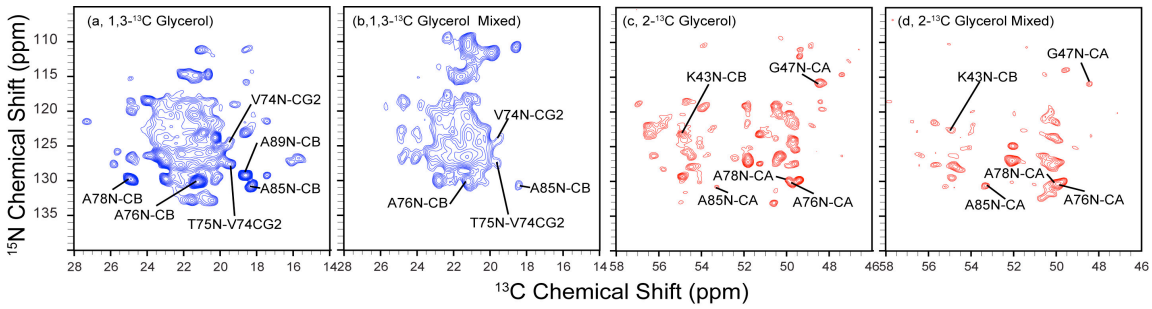

Figure 5.7: 2D TEDOR plans of wt AS fibrils.

2D planes from (a) uniformly 1,3-^{13}C glycerol ^{15}N labeled and (b) 50:50 1,3-^{13}C glycerol, natural abundance (n.a.) nitrogen: n.a. carbon, ^{15}N labeled AS and (c) uniformly 2-^{13}C glycerol ^{15}N labeled and (d) 50:50 2-^{13}C glycerol, n.a. nitrogen: n.a. carbon, ^{15}N labeled AS. TEDOR mixing times of (a,c) 5.76 ms and (b,d) 14.4 ms were chosen to maximize sequential, and long-range correlations respectively.

TEDOR experiments with 14.4 ms of NC mixing on mixed labeled samples yielded 2D planes (Figure 5.7 b,d) with remarkably similar appearance to those of undiluted AS with 5.76 ms of mixing (Figure 5.7 a,c). At 500 MHz and 11.111 KHz spinning a TEDOR mixing time of 14.4 ms will yield long-range correlations, while 5.76 ms mixing will yield the strongest 2 to 3 Å correlations. Therefore the similarity of these spectra supports the hypothesis that a majority of residues are within parallel in-register beta-sheets, where the cross-sheet, intermolecular contacts would have identical chemical shifts to the sequential correlations. Due to the partial degeneracy of the amide ^{15}N chemical shifts, and the small chemical shift dispersion of the C', CA and Cmethyl ^{13}C signals, the majority of cross peaks in such a 2D spectrum have ambiguous assignments. Among the uniquely resolved and unambiguously assigned cross peaks, A89N-CA and A85N-CA (Figure 5.7a,b) as well as A85N-CB and V74N-CG2, (Figure 5.7c,d) are consistent

with a parallel, in-register arrangement of AS. Many other peaks are ambiguous but could include a sequential assignment. Some peaks are notably missing, including A78N-CB and A89N-CB indicating these side chains many be interacting more closely with intramolecular contacts.

5.7 Semi-Automated Construction of High-Resolution Models with Distance Geometry and Simulated Annealing

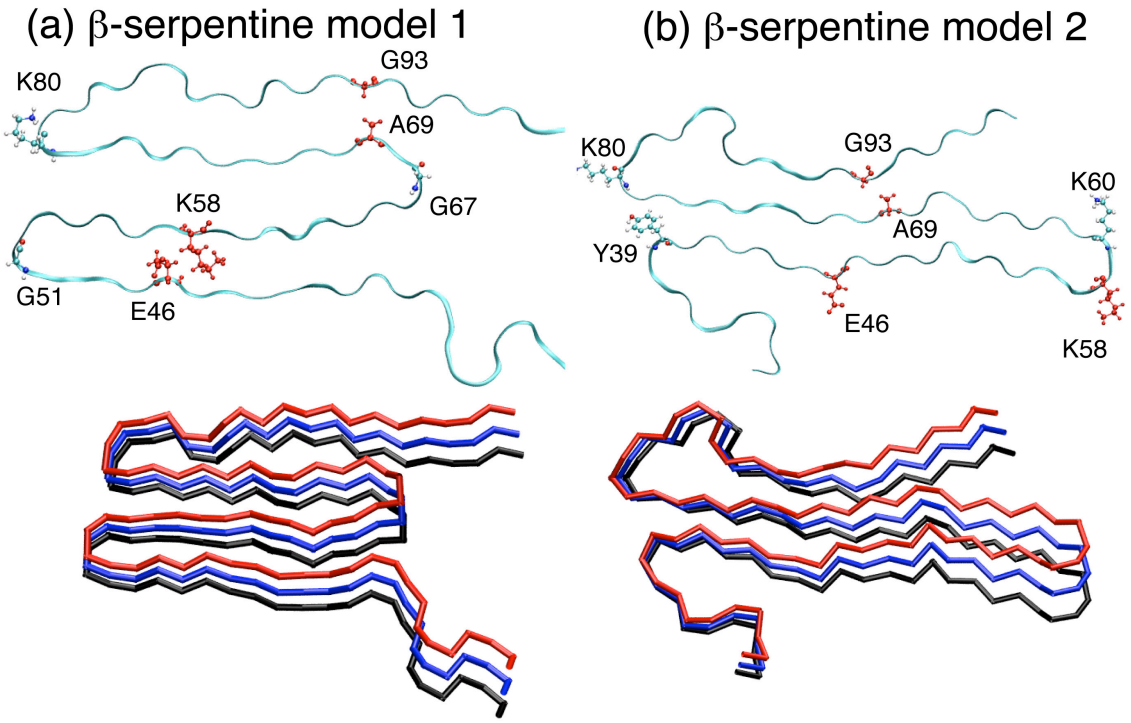

Figure 5.8: Two beta-serpentine models of AS

(a) Beta-serpentine model 1 satisfies the 4.5 Å width requirements while (b) beta-serpentine model 2 more closely follows the 12-residue motif observed in the SSNMR data.

The AS fibril models shown in Figure 5.8 were constructed with X-PLOR-NIH[32] from a combination of experimental and hypothetical data, (a) the beta-serpentine model shown in Figure 5.3 as well as (b) a second beta-serpentine model which adheres more

literally to the 12 residue – 5 residue beta-sheet motif. In addition to the regular bond and angle energies, dihedral restraints from TALOS+ and intermolecular distances based on a parallel in-register beta sheet interpretation of the intermolecular TEDOR experiments were included. The locations of turns and sheets were based on the TALOS+ data and EM and possible alignments of the sheets were considered.

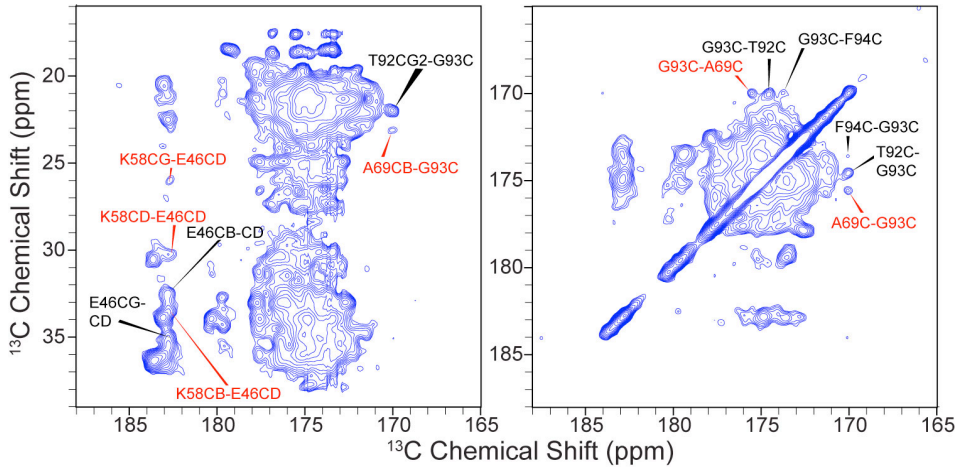

Figure 5.9: 2D ^{13}C-^{13}C DARR spectrum of diluted wt AS.

Sequential correlations are labeled in black while long-range correlations are in red.

For the first beta-serpentine model, two sets of correlations observed in the diluted ^{13}C-^{13}C correlation spectra shown in Figure 5.9 were utilized as anchors for the beta-sheets. Because this sample of AS was fibrilized from a 1:3 mixture of uniformly ^{13}C,^{15}N labeled AS and n.a. AS, intermolecular correlations are greatly attenuated and peaks observed are assumed to be intramolecular. The first set consists of correlations between the completely resolved G93C', to the C' and CB carbons of A69 (a G93CA-A69CB peak exists as well but is not shown). The second is the interaction between the well-resolved CD of E46 to the side chain of K58. Both sets of residues are colored red in Figure 5.8. These two sets of correlations help to propose an alignment of sheets that

correlate with the 12-residue motif and match the EM width. From these alignments hypothetical intramolecular CA-CA distances were generated. Finally, each monomer was forced to be identical by a pairwise non-crystallographic symmetry restraint. In later calculations, ambiguously assigned intramolecular data taken from CC and CNhH data sets, discussed in Chapter 4, were also applied. Consistent with the non-crystallographic symmetry, the ambiguity refers to an uncertainty about which amino acid residue is observed, not the chain. The requirement that each sheet be ~12 residues long matched nicely with the 12 residue sheet segments observed in the TALOS+, yet necessitated that one of the 5 residue sheet – loop – 5 residue sheet segments be utilized as a kinked ~12 residue sheet segment. This segment, centered on the loop at K60, is thus suspect. One solution for this discrepancy is beta-serpentine model 2 (Figure 5.8b), which extends the 2nd and 3rd sheets with out to residue 60, thus more faithfully reproducing the TALOS+ results at the expense of a monomer slightly longer than expected from EM data. Correlations from A69 to G93 are preserved in this model, while E46 to K58 correlations are not. One potential explanation for this would be that E46 could be interacting with one of the lysine residues in the N-terminal domain, although correlations to the most likely candidate, neighboring K45, appear only at the CB and CA positions.

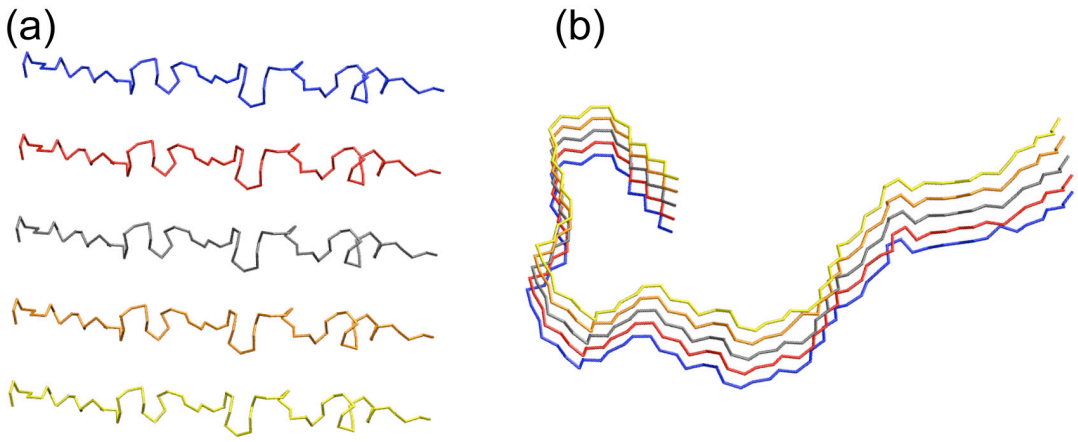

Figure 5.10: Initial structures in AS model production.

(a) The initial conditions for structure calculations show 5 AS monomers (residues 29 to 100) separated by ~20 Å. (b) Calculations performed with only intermolecular distance restraints and TALOS are an intermediate to other beta-serpentine structures.

The initial condition for the calculations consisted of 5 AS monomers (residues 29 to 100) in a loose extended structure and is shown in Figure 5.10a. The monomers were arranged in parallel separated by about 20 Å. Intermolecular restraints were applied to each neighbor (A to B, B to C, C to D etc) and dihedrals and intramolecular distances repeated for each monomer. Because the intermolecular restraints enforce a parallel in-register orientation, and the starting condition is very favorable to that outcome; convergence of the intermolecular packing was good. Control calculations without intramolecular restraints demonstrated that the intermolecular beta-sheets form with just the TALOS+ and dihedral restraints as shown in Figure 5.10b. As a result, any set of self-consistent intramolecular distances allowed for a particular beta-serpentine fold to be modeled. Once models are constructed, the inclusion of experimental data in the

calculation allows a model to be tested against ambiguous distances based on SSNMR data.

5.8 Fully Automated De Novo PASD Calculations of AS Fibrils

A major problem with semi-automated model building of fibril structures became apparent in the course of constructing many different beta-serpentine models of AS: ambiguous intramolecular distances were not sufficient to resolve the differences among the potential models. We attribute this to the highly repetitive sequence of AS, in combination with a high degree of chemical shift degeneracy. Thus, despite a large quantity of high quality experimental data, the distance restraints were unable to distinguish among several plausible beta-serpentine folds. One shortcoming of the XPLOR-NIH ambiguous distance feature complicated the computation of structures: XPLOR-NIH utilizes a potential function in which all restraints are always active and a given restraint is satisfied if any of the possible assignments are satisfied. In order to effectively sample all the possible assignments, it was therefore necessary to utilize more advanced methods. The PASD algorithm[33] within X-PLOR-NIH addresses this issue by using a Monte Carlo assignment strategy to assign unassigned peaks in NOE data sets of proteins with known chemical shift assignments. PASD checks the likelihood of a given peak assignment against intermediate structures, activating and deactivating NOEs using a Monte Carlo algorithm to avoid local minimums in conformational space. After each iteration of the structure calculation, a peak is given a potential assignment and a probability that assignment is correct. Additional runs can be performed until as many peaks as possible have confident assignments. At the

conclusion of the PASD calculation phase, peak lists can be converted to standard format XPLOR distance restraints and used further XPLOR calculations.

The first PASD calculation I performed included the intermolecular TEDOR, TALOS+ dihedrals and peak lists from diluted CC 2D, to be matched to the chemical shift lists generated by Comellas *et. al*.[2] These 2D spectra (one example shown in Figure 5.9) were acquired on the 750 MHz wide bore with 400 ms of DARR mixing on samples of U-^{13}C and 2-^{13}C-glycerol AS, diluted 1:3 in n.a. AS prior to fibrilization. As such, all intermolecular peaks were attenuated by a factor of 4 relative to the intramolecular peaks, giving reasonable confidence the weak long-range peaks observed were intramolecular correlations. The PASD algorithm is also capable of dealing with spurious peaks, either due to noise or intermolecular contacts or experimental artifacts. I therefore included all of the peaks in the CC 2D spectra as intramolecular correlations with which the PASD was performed. In collaboration with Dr. Charles Schwieters (NIH) and Dr. Ming Tang (UIUC), the PASD routines were modified for solid state NMR data sets. The modifications focused on optimizing chemical shift guessing windows, to match the solids linewidths, and, for later calculations, connectivity maps were updated to allow for 3D experiments with NCC dimensions, rather than the CNH, CHH or NHH type 3D spectra more commonly used in solution. Guessing windows were adjusted in order to produce assignment guesses for ~75% of peaks with the wide window and ~50% for the tighter window, mimicking the percentages characteristic of successful solution NMR PASD runs. This was achieved with a wide window of 0.4 ppm and a tight window of 0.2 ppm, relative to an experimental linewidth of 0.3 ppm.

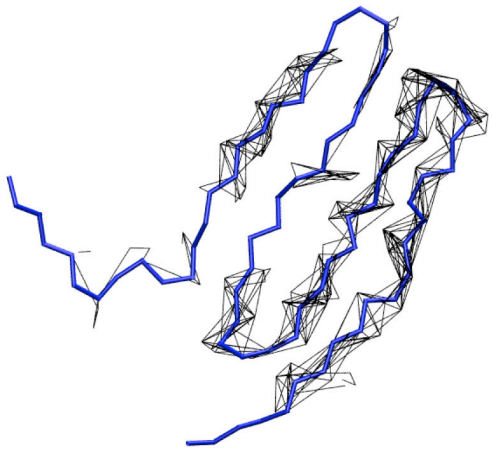

Figure 5.11 Restraints determined through unguided PASD on the AS 5mer.

Distance restraints from diluted CC 2D spectra determined through PASD passes with no additional intramolecular restraints.

To improve the convergence properties of the PASD calculations, the protocol was modified further. Instead of employing only a single Monte Carlo pass, multiple iterations were performed. Following the initial matching iteration, during which the potential assignments for each peak were determined, the Monte Carlo pass was run several times, producing after each run a list of likely assignments and probabilities for each peak. By completing several iterations of the Monte Carlo process, peak assignment likelihoods increased and a final set of restraints was obtained. The distance restraints resulting from this round of calculations consisted entirely of sequential correlations, which can be seen in Figure 5.11 displayed on beta-serpentine model 1 for ease of viewing. The strongest peaks in the 2D's are certainly these sequential peaks, due to proximity in space, and are the easiest for the PASD algorithm to identify unambiguously. This result leaves two main possibilities: either there may be no long-range peaks in these data sets or the long-range peaks may be too degenerate

to allow for convergence. Given our experience with the distance ranges observed in DARR mixing spectra the former is unlikely,[13] so the latter possibility was tested further.

5.9 Validation of Models through PASD Assisted Peak Assignment

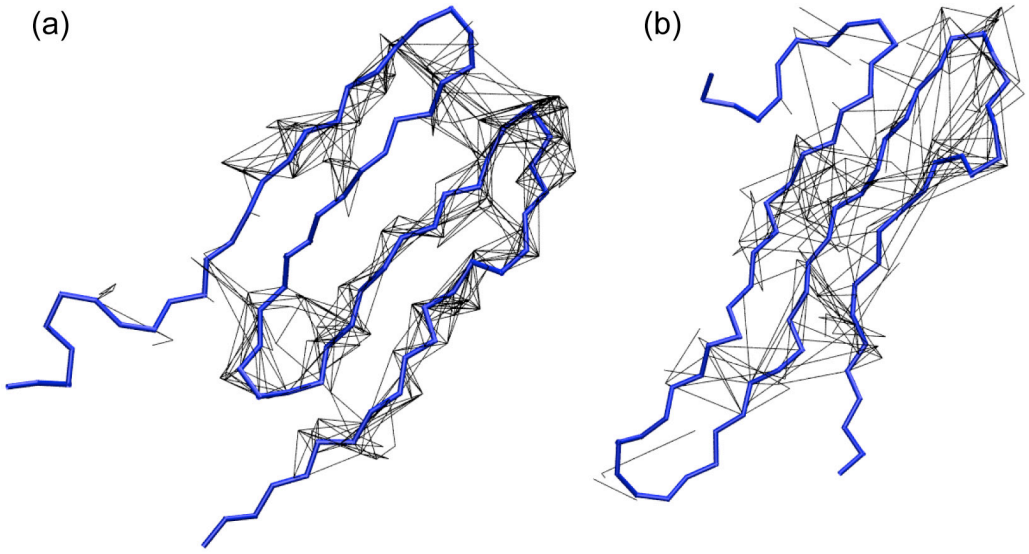

Figure 5.12: Restraints determined through guided PASD calculations.

Distances extracted from diluted CC spectra with restraints enforcing (a) beta-serpentine model 1 and (b) beta-serpentine model 2.

In order to test whether the lack of long-range correlations determined by PASD was due to insufficient convergence issues, two sets of calculations were performed. The first started with the two beta-serpentine models and was performed with the same restraints as the initial PASD calculations. The second started with the models and included hypothetical restraints to drive the formation of model structures as described in Section 5.6. The results from the first set of calculations were basically indistinguishable from the calculations described in the previous section (Section 5.7, Figure 5.11). In contrast, the second set of calculations—performed with the

hypothetical restraints to aid convergence—showed dramatically different results. As illustrated in Figure 5.7a, following successive runs with the hypothetical restraints, PASD was able to converge on a set of long-range peaks in the ^{13}C-^{13}C 2D data that agreed with beta-serpentine model 1. Unfortunately this was also true for the similar beta-serpentine model 2, as shown in Figure 5.12b. The conclusion from these sets of PASD calculations is that the ^{13}C-^{13}C 2D data alone are too degenerate to distinguish between the beta-serpentine models, let alone converge to one structure *de novo*.

The next step therefore was to include 3D experimental data in a manner suitable for PASD computations. 3D experiments were performed as described in Sections 5.2 and 4.5. Two classes of 3D distance experiments were most useful for determination of long-range distances in wt AS fibrils: NCC 3D spectra with ^{13}C-^{13}C DARR mixing between the second and third dimension, and CNhH 3D spectra with ^{1}H-^{1}H RFDR mixing (also between the second and third dimension). The 3D NCC experiment was performed with high digitization and signal averaging: an NCACX 3D acquired at 600 MHz on a U-^{15}N, 2-^{13}C-glycerol labeled AS sample with 500 ms of DARR mixing (averaged for 5 days), while the CNhH was a CANhH with 8 ms of RFDR mixing acquired at 750 MHz on a U-$^{13}C^{2}H^{15}N$ labeled AS sample with the amide sites back-exchanged with 100% $^{1}H_2O$. In both cases, the CA region was used for chemical shift resolution since it provides larger dispersion than the CO region. In the 2D NCA plane inherent to both experiments, approximately two thirds of the peaks are uniquely resolved. In each experiment, there are two chemical shift dimensions prior to the mixing (the NCA or CAN plane) digitized in the indirect dimension, and one chemical shift dimension after the mixing digitized in the direct dimension. For a 3D, this is the

optimal situation as the two chemical shift dimensions help counter any degeneracy due to broadening that may result from shorter acquisition in the indirect dimensions. We modified PASD, from the default parameters used for solution NMR 3D HMQC-NOESY experiments, to accommodate NCC and CNhH 3D experiments; as noted above, this required modification of the linewidth parameters as well as naming conventions and other bookkeeping procedures. PASD calculations were again performed with the intermolecular TEDOR and TALOS+ dihedral angles, along with the standard energies and NCS symmetry restraints. The results of these calculations again showed that even with 3D data, our ambiguity was still too high to allow for convergence. This is likely because resolution of the peaks is limited by the single chemical shift in the direct dimension, in direct analogy to the 2D CC data sets analyzed first. While there are some sites resolved in one dimension in both a 2-^{13}C glycerol ^{13}C carbon 1D spectrum and a ^{1}H 1D spectrum, apparently these do not provide sufficient unambiguous distances to push the system to convergence.

5.10 Ongoing Experiments and Calculations

The remaining requirement to converge to a single fold of the monomer is to identify unambiguous, unique long-range distances. Two promising solutions to this issue have been presented in the literature. The first accomplishes the resolution of long-range distances through the use of Cys point mutations, followed by the addition of paramagnetic labels.[34] The second through the use of ^{1}H detected 4D experiments.[35] In the paramagnetic relaxation enhancement (PRE) experiments, the position of the paramagnetic is known with good precision, since only one residue is modified to Cys (AS has no Cys residues in its native sequence), and the modified sidechain has few

degrees of freedom. The nuclei in proximity to the spin label can be identified with 2D or 3D experiments, in most cases identifying sites uniquely as a result of previously determined chemical shift assignments. The distance from the spin label to the nucleus is evident in enhanced relaxation. Large transverse PREs for nuclei within ~10 Å of the spin label are observed qualitatively, in that peaks are not observed at all in a paramagnetic sample relative to a diamagnetic control. Smaller PREs are measured by a quantitative determination of the T_1 relaxation in 2D or 3D experiments conducted both on the diamagnetic and paramagnetic samples. These experiments are ongoing in collaboration with Marcus Tuttle.

For the 4D proton detected experiments, resolution is achieved by the addition of the fourth chemical shift dimension, an HXhhXH 4D experiment[35] consists of two, 2D HX planes, with a ^{1}H-^{1}H RFDR or DREAM type mixing period in between. While in principle any 3D distance determination experiment could add a fourth dimension, the increased sensitivity proton detected experiments, due to the high ^{1}H gyromagnectic ratio and fast recycle delays discussed in Chapter 4, are a promising option. We have prepared suitably labeled samples of AS fibrils to conduct the 4D experiment, and further experiments are ongoing at 750 MHz.

5.11 Conclusions

Enormous progress has been made towards determining the structure of wt AS fibrils. Complete chemical shift assignments provided a detailed look at the secondary structure of the fibrils and formed the basis for several plausible models of the structure. These models were tested against distance restraints obtained from uniformly labeled samples for maximum sensitivity and diluted samples optimized for either intermolecular

or intramolecular information. To help counter the high degree of degeneracy in AS spectra, glycerol based ^{13}C-labeling schemes and 3D NMR pulse sequences were employed. In addition, the PASD algorithm from XPLOR-NIH was used to help extract distances from ambitious restraints. To refine our structural models to atomic resolution will require further unambiguous distance information, which PRE and 4D may be able to provide.

5.12 References

1. Kloepper, K. D.; Zhou, D. H.; Li, Y.; Winter, K. A.; George, J. M.; Rienstra, C. M., "Temperature-dependent sensitivity enhancement of solid-state NMR spectra of alpha-synuclein fibrils." *J. Biomol. NMR* **2007,** *39* (3), 197-211.
2. Comellas, G.; Lemkau, L. R.; Nieuwkoop, A. J.; Kloepper, K. D.; Ladror, D. T.; Ebisu, R.; Woods, W. S.; Lipton, A. S.; George, J. M.; Rienstra, C. M., "Structured Regions of alpha-Synuclein Fibrils Include the Early-Onset Parkinson's Disease Mutation Sites." *J. Mol. Biol.* **2011**.
3. Spillantini, M. G.; Schmidt, M. L.; Lee, V. M. Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M., "alpha-Synuclein in Lewy bodies." *Nature* **1997,** *388* (6645), 839-840.
4. Polymeropoulos, M. H.; Lavedan, C.; Leroy, E.; Ide, S. E.; Dehejia, A.; Dutra, A.; Pike, B.; Root, H.; Rubenstein, J.; Boyer, R.; Stenroos, E. S.; Chandrasekharappa, S.; Athanassiadou, A.; Papapetropoulos, T.; Johnson, W. G.; Lazzarini, A. M.; Duvoisin, R. C.; Dilorio, G.; Golbe, L. I.; Nussbaum, R. L., "Mutation in the alpha-synuclein gene identified in families with Parkinson's disease." *Science* **1997,** *276* (5321), 2045-2047.
5. Kruger, R.; Kuhn, W.; Muller, T.; Woitalla, D.; Graeber, M.; Kosel, S.; Przuntek, H.; Epplen, J. T.; Schols, L.; Riess, O., "Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease." *Nat. Genet.* **1998,** *18* (2), 106-108.
6. Singleton, A. B.; Farrer, M.; Johnson, J.; Singleton, A.; Hague, S.; Kachergus, J.; Hulihan, M.; Peuralinna, T.; Dutra, A.; Nussbaum, R.; Lincoln, S.; Crawley, A.; Hanson, M.; Maraganore, D.; Adler, C.; Cookson, M. R.; Muenter, M.; Baptista, M.; Miller, D.; Blancato, J.; Hardy, J.; Gwinn-Hardy, K., "alpha-synuclein locus triplication causes Parkinson's disease." *Science* **2003,** *302* (5646), 841-841.
7. Zarranz, J. J.; Alegre, J.; Gomez-Esteban, J. C.; Lezcano, E.; Ros, R.; Ampuero, I.; Vidal, L.; Hoenicka, J.; Rodriguez, O.; Atares, B.; Llorens, V.; Tortosa, E. G.; del Ser, T.; Munoz, D. G.; de Yebenes, J. G., "The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia." *Ann. Neurol.* **2004,** *55* (2), 164-173.
8. Davie, C. A., "A review of Parkinson's disease." *Br Med Bull* **2008,** *86*, 109-27.
9. Desplats, P.; Lee, H. J.; Bae, E. J.; Patrick, C.; Rockenstein, E.; Crews, L.; Spencer, B.; Masliah, E.; Lee, S. J., "Inclusion formation and neuronal cell death through neuron-to-neuron transmission of alpha-synuclein." *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (31), 13010-13015.

10. Luk, K. C.; Song, C.; O'Brien, P.; Stieber, A.; Branch, J. R.; Brunden, K. R.; Trojanowski, J. Q.; Lee, V. M. Y., "Exogenous alpha-synuclein fibrils seed the formation of Lewy body-like intracellular inclusions in cultured cells." *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (47), 20051-20056.
11. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H., "Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core." *Science* **2008,** *319* (5869), 1523-1526.
12. Ritter, C.; Maddelein, M. L.; Siemer, A. B.; Luhrs, T.; Ernst, M.; Meier, B. H.; Saupe, S. J.; Riek, R., "Correlation of structural elements and infectivity of the HET-s prion." *Nature* **2005,** *435* (7043), 844-848.
13. Franks, W. T.; Wylie, B. J.; Schmidt, H. L. F.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M., "Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (12), 4621-4626.
14. Wylie, B. J.; Schwieters, C. D.; Oldfield, E.; Rienstra, C. M., "Protein Structure Refinement Using 13Cα Chemical Shift Tensors." *J. Am. Chem. Soc.* **2009,** *131* (3), 985-992.
15. Van Melckebeke, H.; Wasmer, C.; Lange, A.; Ab, E.; Loquet, A.; Bockmann, A.; Meier, B. H., "Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state NMR spectroscopy." *J. Am. Chem. Soc.* **2010,** *132* (39), 13765-75.
16. Kloepper, K. D.; Woods, W. S.; Winter, K. A.; George, J. M.; Rienstra, C. M., "Preparation of alpha-synuclein fibrils for solid-state NMR: Expression, purification, and incubation of wild-type and mutant forms." *Protein Expres. Purif.* **2006,** *48* (1), 112-117.
17. Himber, J.; Refino, C. J.; Burcklen, L.; Roux, S.; Kirchhofer, D., "Inhibition of arterial thrombosis by a soluble tissue factor mutant and active site-blocked factors IXa and Xa in the guinea pig." *Thromb. Haemost.* **2001,** *85* (3), 475-481.
18. Heise, H.; Hoyer, W.; Subramaniam, V.; Jovin, T. M.; Baldus, M. In *Solid-state NMR Investigation of Uniformly 13C,15N Labeled alpha-Synuclein Fibrils*, Experimental NMR Conference, Experimental NMR Conference, Asilomar, CA, Experimental NMR Conference, Asilomar, CA, 2004.
19. Loquet, A.; Giller, K.; Becker, S.; Lange, A., "Supramolecular interactions probed by 13C-13C solid-state NMR spectroscopy." *J. Am. Chem. Soc.* **2010,** *132* (43), 15164-6.
20. Vilar, M.; Chou, H.-T.; Luhrs, T.; Samir, M.; Riek-Loher, D.; Verel, R.; Manning, G.; Stahlberg, H.; Riek, R., "The fold of alpha-synuclein fibrils." *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (25), 8637-8642.
21. Gath, J.; Habenstein, B.; Bousset, L.; Melki, R.; Meier, B. H.; Bockmann, A., "Solid-state NMR sequential assignments of alpha-synuclein." *Biomol NMR Assign* **2011**.
22. Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M., "Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR." *Proc. Natl. Acad. Sci. USA* **2005,** *102* (44), 15871-15876.
23. Der-Sarkissian, A.; Jao, C. C.; Chen, J.; Langen, R., "Structural organization of alpha-synuclein fibrils studied by site-directed spin labeling." *J. Biol. Chem.* **2003,** *278* (39), 37530-37535.

24. Petkova, A. T.; Leapman, R. D.; Guo, Z. H.; Yau, W. M.; Mattson, M. P.; Tycko, R., "Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils." *Science* **2005,** *307* (5707), 262-265.
25. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A., "TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." *J. Biomol. NMR* **2009,** *44* (4), 213-223.
26. Takegoshi, K.; Nakamura, S.; Terao, T., "C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR." *Chem. Phys. Lett.* **2001,** *344* (5-6), 631-637.
27. Lange, A.; Luca, S.; Baldus, M., "Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids." *J. Am. Chem. Soc.* **2002,** *124*, 9704-9705.
28. Jaroniec, C. P.; Filip, C.; Griffin, R. G., "3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon-nitrogen distances in uniformly C-13, N-15-labeled solids." *J. Am. Chem. Soc.* **2002,** *124* (36), 10728-10742.
29. Bennett, A. E.; Ok, J. H.; Griffin, R. G.; Vega, S., "Chemical shift correlation spectroscopy in rotating solids: Radio-frequency dipolar recoupling and longitudinal exchange." *J. Chem. Phys.* **1992,** *96*, 8624-8627.
30. Baldus, M.; Petkova, A. T.; Herzfeld, J.; Griffin, R. G., "Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems." *Mol. Phys.* **1998,** *95* (6), 1197-1207.
31. Sperling, L. J.; Berthold, D. A.; Sasser, T. L.; Jeisy-Scott, V.; Rienstra, C. M., "Assignment strategies for large proteins by magic-angle spinning NMR: the 21-kDa disulfide bond forming enzyme DsbA." *J. Mol. Biol.* **2010,** *399*, 268-282.
32. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., "The Xplor-NIH NMR molecular structure determination package." *J. Magn. Reson.* **2003,** *160* (1), 65-73.
33. Kuszewski, J.; Schwieters, C. D.; Garrett, D. S.; Byrd, R. A.; Tjandra, N.; Clore, G. M., "Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments." *J. Am. Chem. Soc.* **2004,** *126* (20), 6258-6273.
34. Nadaud, P. S.; Helmus, J. J.; Kall, S. L.; Jaroniec, C. P., "Paramagnetic Ions Enable Tuning of Nuclear Relaxation Rates and Provide Long-Range Structural Restraints in Solid-State NMR of Proteins." *J. Am. Chem. Soc.* **2009,** *131* (23), 8108-8120.
35. Huber, M.; Hiller, S.; Schanda, P.; Ernst, M.; Bockmann, A.; Verel, R.; Meier, B. H., "A proton-detected 4D solid-state NMR experiment for protein structure determination." *ChemPhysChem* **2011,** *12* (5), 915-8.